REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE October 9, 1995	3. REPORT TYPE AND DATES COVERED Final Report (08/01/90-05/31/95)

4. TITLE AND SUBTITLE
FRISC - A 1.0 GOPS Fast Reduced Instruction Set Computer for Super Workstations and Parallel Processing Applications

5. FUNDING NUMBERS
DAAL03-90-G-0187

6. AUTHOR(S)
John F. McDonald

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Center for Integrated Electronics
Rensselaer Polytechnic Institute
Troy, New York 12180

8. PERFORMING ORGANIZATION REPORT NUMBER
FRISC-01

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARO 28329.3-EL

11. SUPPLEMENTARY NOTES
The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

19960209 107

13. ABSTRACT (Maximum 200 words)

A Fast Reduced Instruction Set Computer is designed using the Rockwell 50 GHz Baseline GaAs/AlGaAs Heterojunction Bipolar Transistor Process. The goal of the project is a 1000 MIPS integer RISC engine in the form of a multiple chip integration using MCM technology. Cell libraries and CAD tools for this effort have been created and tested. Test structures have been fabricated and tested which indicate that the general approach is viable, and all architecture chips have been completed. However, certain process deficiencies could lead to circuit performance that is slower than predicted from Rockwell supplied models by a factor of 30-50%. In spite of this, "work-around" strategies have been devised and it is felt that the Rockwell process can still achieve the speed goal of 2 GHz clock operation. These strategies consist of altering the layout of the transistor using a smaller emitter with a larger aspect ratio, and three active edges, and some shortening of base-emitter and base collector separations, and variation of the wire geometries or shifting to the third level of metal for them.

14. SUBJECT TERMS
Fast RISC, GaAs/AlGaAs HBT, 1000 MIPS
Subnanosecond computing, Microwave MCM

15. NUMBER OF PAGES
59

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

DTIC QUALITY INSPECTED 1

F-RISC -
A 1.0 GOPS FAST REDUCED INSTRUCTION SET COMPUTER
FOR SUPER WORKSTATION AND TERAOPS PARALLEL
PROCESSOR APPLICATIONS

FINAL CONTRACT REPORT

JOHN F. MCDONALD

JULY 20, 1995

U. S. ARMY RESEARCH OFFICE

INSTITUTION:

RENSSELAER POLYTECHNIC INSTITUTE
CENTER FOR INTEGRATED ELECTRONICS
TROY, NEW YORK 12180
(518)-276-2919
EMAIL:MCDONALD@UNIX.CIE.RPI.EDU

APPROVED FOR PUBLIC RELEASE

DISTRIBUTION UNLIMITED

I.   FOREWORD

Military and commercial systems are becoming increasingly dependent on computers and communication networks for information processing. The speed of digital circuits is a key limitation for these systems. Therefore it is of the utmost importance that the United States possess the technological infrastructure to insert the highest performance devices in critical systems to maintain its leadership edge both in economic and foreign policy endeavors. Although procurement emphasis for military and nonmilitary systems is increasingly placed on Commercial-Off-The-Shelf (COTS) components for cost effectiveness, it is imperative that this philosophy not limit our vision regarding what is possible using more advanced technology. Some future adversary might discover, develop, and master alternate technologies over a period of time. These could prove sufficiently effective to change the balance of power. The Fast Reduced Instruction Set Computer (F-RISC) project has been undertaken to explore the highest speed possible for computer clock rates using some of the most advanced devices that have been developed in the US. The project has capitalized on existing GaAs/AlGaAs Heterojunction Bipolar Transistors (HBT's) and microwave compatible Multichip Modules (MCM's) as the vehicles to achieve these goals. The project can be expected to impact applications ranging from "super" workstations, and parallel processing nodes in TeraOPS computers, to virtual reality engines for simulation, media access controllers for fast microwave communication networks, and direct Digital Signal Processing (DSP) at high frequencies. These latter applications might be suitable for radar, high speed encryption/decryption, and data compression/decompression.

The goal established for this first ARPA/ARO grant of the F-RISC series has been to create a demonstration Fast RISC integer engine with a 2 GHz clock rate and a peak throughput of 1,000 MIPS. Rockwell International offered the Rensselaer team the opportunity to employ their 50 GHz baseline HBT process for this project. Typical gate delays for that HBT process were revealed by Rockwell to be approximately 25 picoseconds, and with reasonable pipelining it has been possible to create an architecture that could respond in about 10 gate delays per clock phase, or 250 picoseconds. Given the low initial yield expected with this process a multichip architecture rather than a monolithic single chip microprocessor was proposed. Typical chip yields of 20% at 5,000 HBT's were assumed for the purpose of the demonstration originally, but this needed to be upgraded to 8,000 HBT's during the course of the project. Most of the additional devices were needed to make the chips testable at microwave frequencies using boundary scan based, embedded at-speed test circuitry. Fortunately, Rockwell's yields improved during the period of this project to meet this requirement.

This first goal of the program paves the way toward other, still higher clock rate systems that could be created in the future. For example, during the period of this work it became clear that a yield improvement for the 50 GHz baseline process to 30,000 HBT's could create the opportunity to double the speed of the system to 2,000 MIPS with some minor architectural changes. Furthermore, Rockwell revealed that a 100 GHz upgrade to the 50 GHz baseline process might make another clock doubling possible to achieve 4,000 MIPS. A superscalar upgrade of the design might then achieve 8,000 MIPS. Finally, the existence of still faster HBT's, up to 320 GHz, suitable for digital design were disclosed to the Rensselaer team, suggesting that 3-4 times higher speeds will eventually become feasible. Because these speeds are well above any projections for CMOS in the SIA

roadmap the Rensselaer team selected Rockwell as an industrial partner for the F-RISC project.

To date the project has accomplished nearly all of its goals. An integrated circuit HBT cell library has been developed, CAD tools unique for the project requirements have been developed and tested, the four architecture chips for F-RISC have been designed, and checked extensively, and finally test circuits have been fabricated to help verify process, device and circuit models. The four architecture integrated circuits are to be fabricated on funds still associated with the budget for this project and which have been committed to Rockwell through a purchase order.

Challenges emerged during the project as speed discrepancies were discovered between the original HBT SPICE models supplied by Rockwell and measured transistor performance in fabricated test structures. Additional discrepancies were discovered regarding thickness of the polyimide interlevel dielectric (ILD) in different circuit regions on our test chips. This latter problem was discovered on chips fabricated under companion funding, subcontracted to us by Rockwell under the HSCD BAA. With Rockwell's collaboration we are currently investigating device speed improvements with smaller emitter areas and scaling interconnections to address these challenges.

A follow on contract work has already been awarded under the HPCS BAA which concentrates on solving the speed problem, creating device and interconnect layouts that compensate for the device modeling error, and which fabricates demonstration architecture chips. Solutions for regaining this speed are being sought in a manner that permits use of the existing architecture chips with only simple transistor substitutions and interconnection transformations; a strategy which thereby preserves most of the investment in the architecture from this contract. In addition the follow-on work will continue to fabricate chips till a sufficient number of Known Good Die (KGD) are available to populate several MCM prototypes, and design the MCM layout. At that point funding would be required to insert these chips into an MCM to build a Fast RISC module. A proposal for this work has been submitted under BAA 95-26, for mixed mode MCM's. That proposal has been assigned a status of "selectable," subject presumably to satisfactory performance under the present HPCS BAA and availability of funding.

Table of Contents

I.1. List of Figures

II. FINAL REPORT

II.1. STATEMENT OF THE PROBLEM STUDIED

For the past two decades the speed of computers and communication networks has increasingly been dictated by circuits implemented in commercially attractive Complementary MOS or CMOS digital circuit technology. CMOS has exhibited a long trend of providing higher performance computation and communication systems at lower and lower prices. However, there are some disturbing indications that this trend will not continue, at least not at the same pace. Notably, the cost of fabrication facilities for this technology is increasing dramatically. This is due in some part to the cost of lithography for the smaller circuit features needed to attain still higher levels of performance. Additionally certain fundamental device, process, and circuit limitations are emerging for these smaller devices which could end the trend exemplified by Moore's law. Moore's law predicts a doubling in circuit performance every three years. Industry has come to depend upon this trend that makes computer hardware obsolete after 3-6 years and drives customers to upgrade their hardware in the same time frame. Recently, some published articles on industry trends have brought attention to the fact that this trend is slowing down to one doubling factor every four to five years. Several factors have contributed to this slow down, some of these represent permanent paradigm shifts.

One of these shifts is due to changes in the importance of interconnections in integrated circuits. Increasingly interconnections dominate system speed. This is due to the emerging importance of the resistance of these connections because of their reduced cross sectional area. Voltage scaling of devices limits the supply voltage to about 1.5-2.0 Volts. Short channel effects make it difficult to maintain turn-on/turn-off characteristics for these devices, and their ability to drive interconnections grows weaker. This is why even a successful deep submicron device technology may have difficulty showing a performance improvement in real systems by simple scaling to smaller dimensions. More importantly, even if such performance could be realized there is a severe question of the cost associated with manufacturing such small devices.

The computing engines created in CMOS have increased in *architectural complexity*, exploiting parallelism and pipelining implicit in some algorithms. The complexity has increased the difficulty of *design*, and makes the development of new computer architectures much more difficult and costly. These are significant cost factors that are often overlooked in projecting future trends of system costs. Moreover, there is concern that CMOS may not provide the fastest technology for digital circuits. If an adversary were to establish a foothold in a superior alternative circuit technology it could significantly alter the future balance of economic and military power. Consequently it is prudent to carefully evaluate whether alternatives exist which would permit the construction of faster computers though the use of faster devices that scale differently, or at least more forgivingly, compared with CMOS.

II.1.A. The search for superior alternate devices and technologies

The Rensselaer Fast RISC project was created to explore alternative devices and materials systems which present the opportunity to create circuits that could ultimately outperform CMOS in digital computers. This search led to the selection of the Heterojunction Bipolar Transistor (HBT) in the GaAs/AlGaAs materials system as the best

1

starting candidate for this project. It represented the most advanced III-V technology available during the time period of 1990 when the contract started. At the time of the writing of this report it is still the fastest device technology available to us. The HBT fabrication facility is termed a 50 GHz baseline process because the device exhibits a 50 GHz transit frequency at optimal collector current and collector emitter voltage. The peak transit time frequency is the inverse of the time required for an electron to traverse the base region of an n-p-n HBT. This time is defined roughly to satisfy the notion that while an electron passes through the device control region, the field it sees must remain relatively constant. That way the current passing through the base can still track the voltage applied to the base. Baring the effects of other circuit parasitics and second order effects, this peak transit frequency establishes the highest frequency that circuits can realize using the devices in that process.

A 50 GHz peak transit time frequency device SPICE model supplied by Rockwell suggested this device might be appropriate for realization of a demonstration 1000 MIPS (2.0 GHz four phase clock) machine. Unloaded inverter delays of 15-18 picoseconds were predicted by this model. Loaded inverter delays of 25 ps were predicted for a high power gate with 100 fF of wire loading. One can argue that with proper pipelining and packaging it should be possible to implement a RISC engine in roughly 20 gate delays, the time for an accelerated addition, or approximately the time for one register file access. Furthermore, future versions of this device appear promising for the realization of even faster machines. It is believed that existing HBT technology is capable of providing digital circuit speeds far in excess of those possible in CMOS. The question is whether this technology can be used cost effectively for fast computing nodes, super workstations, and ultra fast networks. The total investment in developing alternative technologies is low compared to the investments in facilities for fabrication of deeply submicron CMOS technology. The steady rate of progress in CMOS represents a challenge to the introduction of alternative technologies. The advantages of alternative technologies relative to CMOS must be large enough to warrant the commitment of funds. Thus alternative technologies need first to demonstrate device yields sufficient for commercial and military computing as well as signal processing applications to open sufficiently large revenue streams to allow them to aggressively push process development towards higher integration levels which would lower costs and increase the range of applications considerably.

Although the modern trend in processor design is towards Multiple Parallel Processors (MPP's) and Networks of Workstations (NOW's) these architectures tend to be slow when the type of algorithms run on them does not lend itself to parallelization, or demands excessive interprocessor communication. By making the processors faster, fewer of them are required for a TERAOPS system. Another benefit from having fewer nodes is that it cuts down on interprocessor communications required for task or thread synchronization. If this higher speed is attained at similar levels of power dissipation per MIPS to CMOS, then a better computing environment is obtained, one which is easier to program.

In selecting the HBT other device and process technologies had to be thoroughly evaluated in order to determine whether it was the best choice. The natural competitor for the GaAs/AlGaAs HBT is the MESFET. MOSFET or MESFET technology depends on lithographic shrinkage's for improvements in performance. To excel in speed the transit time of an electronic carrier through the *horizontal* channel of the device must be made short because this horizontal channel is the control region of the device. Electrons have to cross this region in a time fast enough that the gate control voltage appears essentially constant in order to have the source drain current respond to the gate voltage. Hence, high speed must be obtained by shrinking the gate length or increasing the carrier velocity. In addition, to shortening the gate length the integrated circuit interconnections must also scale in length to achieve higher speed. The usual argument in favor of GaAs technology is that

the carrier velocity will be higher for electrons due to higher electron mobility (at low field strength), and hence higher speed should be seen for a given gate length with GaAs MESFETS. However, there are other factors, such as wire loading effects which can mask these advantages. If the MESFET technology cannot provide the number of interconnect levels and minimal interconnect geometries provided by advanced CMOS processes to keep wirelengths down, then the advantages expected from the higher mobility may not be seen at the circuit level.

By comparison the Rensselaer effort has focused on bipolar devices. In vertical bipolar npn transistor technology the most fundamental speed limitation is determined primarily by the thickness of epitaxial material layers, especially the thin base, which is the _vertical_ control region for this device. There is a secondary dependency on the horizontal dimensions on horizontal lithographic dimensions. This secondary dependency should not be construed to imply, however, that horizontal dimensions are unimportant for the bipolar device. We shall see that these secondary considerations cannot be ignored. However, it is generally true that for a given lithographic dimension, with a suitably thin base, the HBT will outperform the MESFET with the same minimum lithographic feature size. Base thicknesses for HBT's are typically 100 nanometers in the 50 GHz Rockwell baseline process, with the 100 GHz process pushing 50 nanometers. These vertical dimensions are readily attainable today in production for the device, whereas for CMOS fabrication at comparable horizontal control reqion dimensions one would require routine use of x-ray lithography.

Therefore, the basic hypothesis remains that transit time through a device is its fundamental limit, and to approach this fundamental limit some attention must be paid to horizontal dimensions of the HBT device to reduce its parasitics. To excel the vertical layer dimension must be made small. _What is claimed_ is that the horizontal dimensions _of the HBT_ need not be scaled as aggressively as in CMOS to obtain superior device performance. To accomplish the thin vertical dimensions the device transit layers must be fabricated by one of the epitaxial growth techniques recently developed (i.e. either MBE or OMCVD). The horizontal dimensions also need to be small, but not nearly as small as these vertical dimensions. However, for a fair comparison the wire loading with these two competing device technologies must be comparable. Large devices promote long wire connections, and so once again the fundamental device limits must make some concessions to the application environment (e.g. wiring dimensions and/or numbers of wiring layers) in which they are to be used. Fortunately for this comparison both the HBT and MESFET circuit lines supported 3 levels of metal with comparable wiring geometries.

MOSFET, MESFET or HEMT devices exhibit less ability to drive interconnections than bipolar devices when driven by other FET devices because of low transconductance, and technically FET's should also be less capable of dealing with the large currents needed to charge and discharge wires rapidly. Peak current flows in FET devices in thin channel regions that are only several tens of nanometers thick in aggressively scaled devices. To charge and discharge capacitive loads the current density in these channels can be quite high. This can lead to thermal damage, or even dopant redistribution in certain materials systems and with certain dopants.

To confirm the suspicion that MESFET implementations of the same architecture could not perform as well at the HBT version, a companion project funded by IBM and Rockwell was launched. This implementation utilized the same F-RISC architecture studied (as represented by a system netlist) in the HBT implementation. However, this MESFET effort resulted in a _monolithic_ or single chip microprocessor realization rather than a multichip system. This should have given the ultimate wire minimization advantage to the MESFET implementation, but would place severe restrictions on heat dissipation. This companion MESFET design was dubbed F-RISC/I, to distinguish it from the HBT effort, which was named F-RISC/G, and to further categorize still other architecture

3

embedding experiments in the future. The F-RISC/I fabrication was implemented at Rockwell utilizing an 0.7 micron E/D H-MESFET process with single ended Super Buffer FET Logic (SBFL) circuits. H-MESFET is a special variant of MESFET called heteroMESFET The standard cell library for this process was provided by IBM and the layout of the chip was completely generated using only one pass with the CADENCE standard cell router with extensive assistance by CADENCE personnel. Due to time pressure, even the highly ordered register file and adder circuits were implemented with standard cells, which did not take advantage of the regularity inherent in these circuits. The results of this fabrication became known at about the end of the third year of the presently completing contract. Through the use of various test circuits and an HP 500 MHz test system at Yorktown Heights, the chip was found to operate at speeds of at least 160 MHz. The boundary scan circuits were tested and found to operate at a shift-in and shift-out rate as high as 400 Mhz. The power dissipation was 3.8 Watts at 160 MHz.

F-RISC/I had a circuit implementation that employed relatively inefficient standard cell layouts, the register file and ALU are not hand crafted as it is in the HBT F-RISC/G effort. Also the device thresholds actually used in fabrication did not match well the ones assumed in the design phase. So a second study was conducted to estimate the speed of a monolithic MESFET F-RISC with more careful layout and the correct device thresholds. This estimate came to 350-400 MHz, or about one third to one half of the speed of the much larger Rockwell HBT devices with design rules of 1.4 μm.
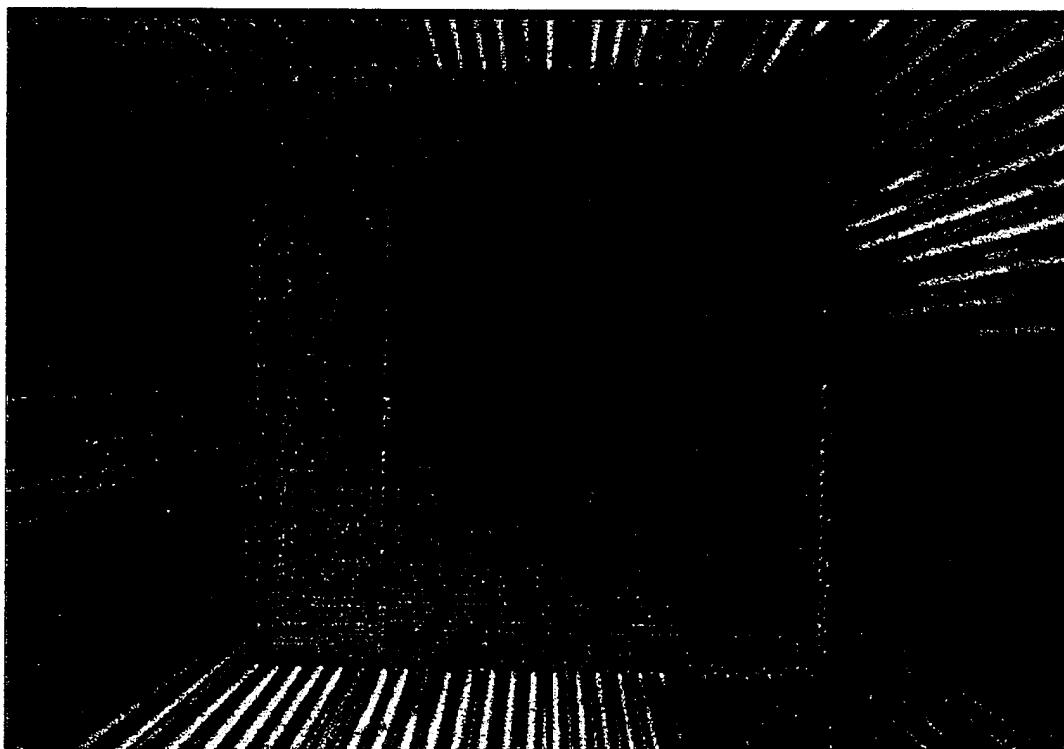
Figure 1. Micrograph of a 0.7 μm monolithic H-MESFET F-RISC/I shown in 400 MHz CERPROBE test card, using the HP test system at Yorktown Heights.

Since clearly higher yields and smaller devices would be possible in the future using the more advanced versions of the HBT process, this helped provide confirmation that the HBT could theoretically provide superior performance, and eventually reach regimes of performance that even deep submicron MESFET or CMOS microprocessors can not reach.

Figure 2 shows the predicted clock frequencies of F-RISC/I implementations for 0.7 μm and 0.5 μm H-MESFET versions as a function of scaled interconnect capacitance. The performance of the 0.5 μm version is based upon the device models provided by Rockwell and the interconnect length is shrunk according to the 0.5 μm process design rules. Clearly the interconnect capacitance has a large effect on the cycle time. A full custom implementation which could reduce interconnect capacitance by about 1/2 would about double the performance of the 0.7 μm and increase the performance of the 0.5 μm version from 350 to 440 MHz.

Another consideration in choosing a device technology is that the collector breakdown voltage of the controlling region must remain high at small thickness. This is important since in predicting future trends the controlling region must inevitably be made thinner. III-V materials systems appear to have a good chance of offering a path to superior switching speed because the product of their peak transit time frequency multiplied by the collector-emitter breakdown voltage exceeds the 230 GHz-Volts physical limit of silicon. For example, a silicon homojunction bipolar transistor with a 60 GHz peak transit time frequency can sustain only about 4 volts by this calculation. To make a faster device would require thinner base regions and the breakdown voltage would be even lower. In GaAs/AlGaAs HBT's a the same peak transit time frequency could sustain 15 Volts, and in InP the breakdown voltage can be 20 V. Certain additional HBT technologies involving SiC and SiGe remain to be explored.

Further considerations relate to the cost of HBT technology, and the power dissipation associated with the circuitry. However, there are only a few key locations where these extremely fast computer and networking circuits are needed. These locations might include network media access controllers for optical or satellite transceivers, direct microwave frequency digital processing, radar signal processing, high frequency data compression/decompression or encryption/decryption or complex nonparallelizable algorithms. In such systems cooling of the processor would not be a problem, and the cost might be acceptable since a CMOS alternative would require a large amount of parallel hardware and introduce very long latencies.
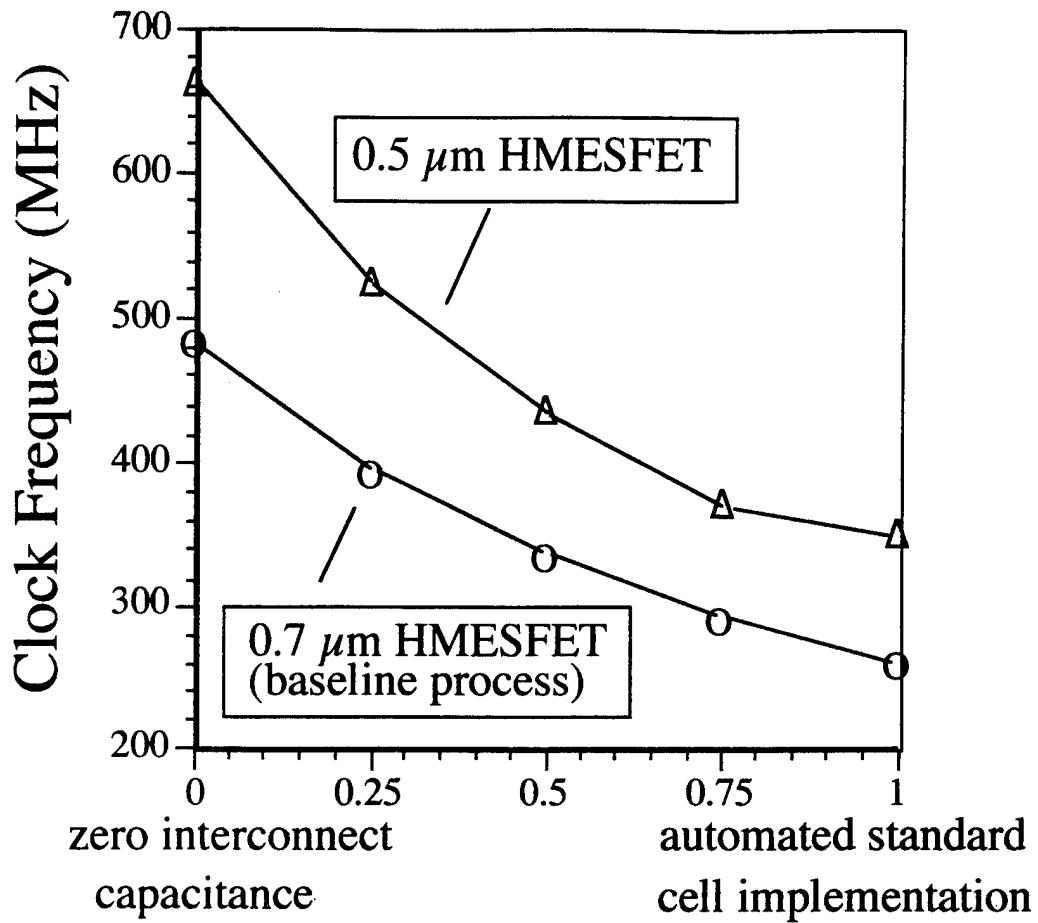
Figure 2. Clock Frequencies of F-RISC/I versions with 0.7 μm and 0.5 μm H-MESFETs as a function of scaled interconnect capacitance.
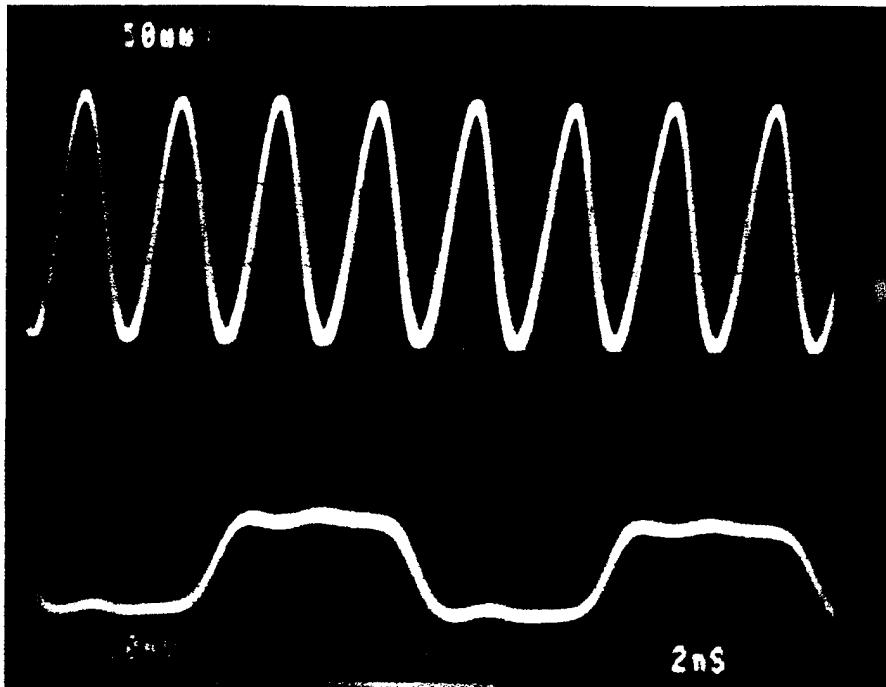
Figure 3. **Sample waveform taken at 400 MHz for the 2X clock of boundary scan activity for testing the 0.7 μm MESFET F-RISC/I at speed.**

To provide the basis for a computer industry, however, HBT devices must make their way into a fabrication process that can provide the capabilities required by LSI or VLSI integrated circuits. For rapid evaluation, our group has limited its attention only to technologies available in commercial production at the startup of the contract. Usually such lines were constructed for other purposes, such as microwave analog applications which require only a very limited number of devices. HBT devices have made their way into digital circuits at very few places. The circuits capable of exhibiting the greatest speed with good HF noise control, namely Current Mode Logic (CML), shown in Figure 5, require three terminal access to the HBT devices. At the inception of the contract only one company offered the Rensselaer group access to such technology in a fabrication line capable of producing circuits containing approximately 5,000 HBT's, namely Rockwell International, located in Newbury Park, California. In Rockwell's case there was a substantial commitment to making both analog and digital circuits. In this way the known success of GaAs/AlGaAs HBT's in analog applications might bolster the existence of the fabrication line. Hence, Rensselaer selected the Rockwell 50 GHz baseline process for its initial experiment in Fast RISC architectures.
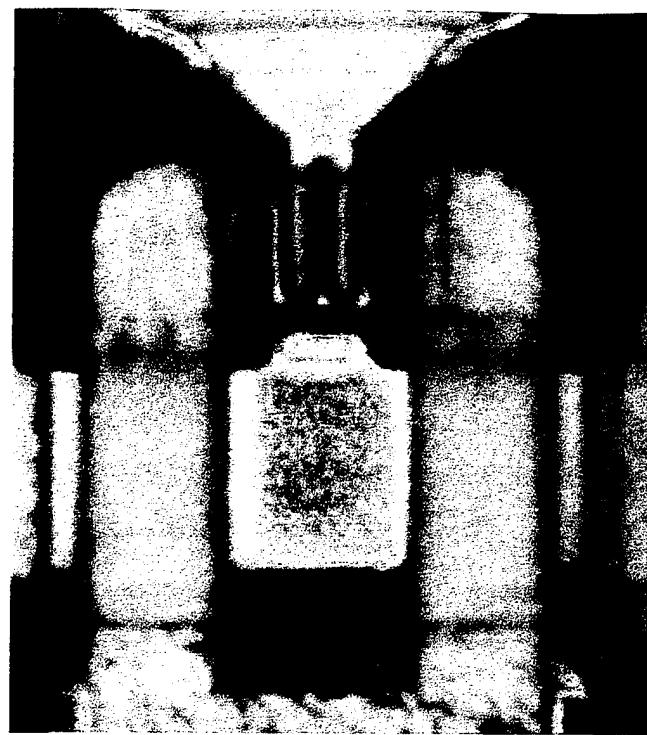
Figure 4. Optical photomicrograph of a 1.4 µm emitter stripe HBT as fabricated in the Rockwell 50 GHz Baseline Process.
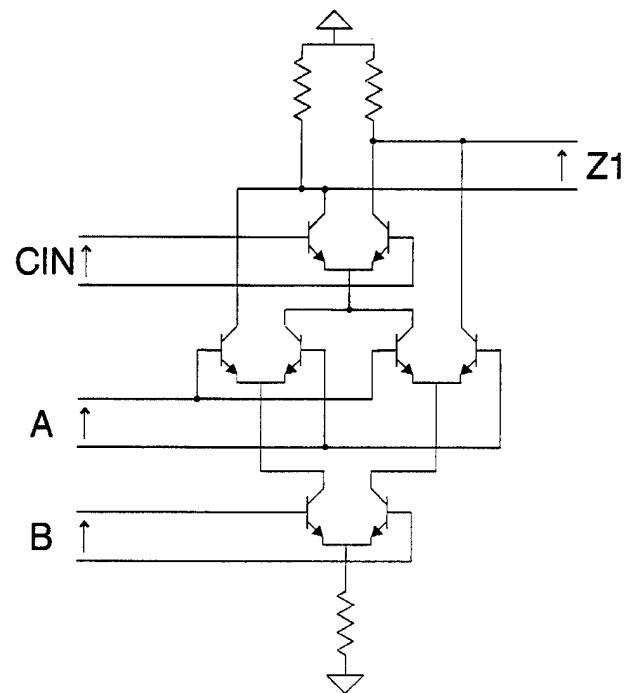
Figure 5. Differential CML gate with three levels of current switches.

The small 5000 HBT yields initially offered by Rockwell, supported only a modest, highly simplified RISC architecture, similar to that of the Berkeley RISC II, with the exception of the large 132 word register file and full 32 bit barrel shifter. Even this modified Berkeley RISC architecture would require a multichip realization with a dense multichip module (MCM) package to reach a 1 ns cycle time. Additionally, the MCM would have to be qualified to support the 2 GHz clock signal.
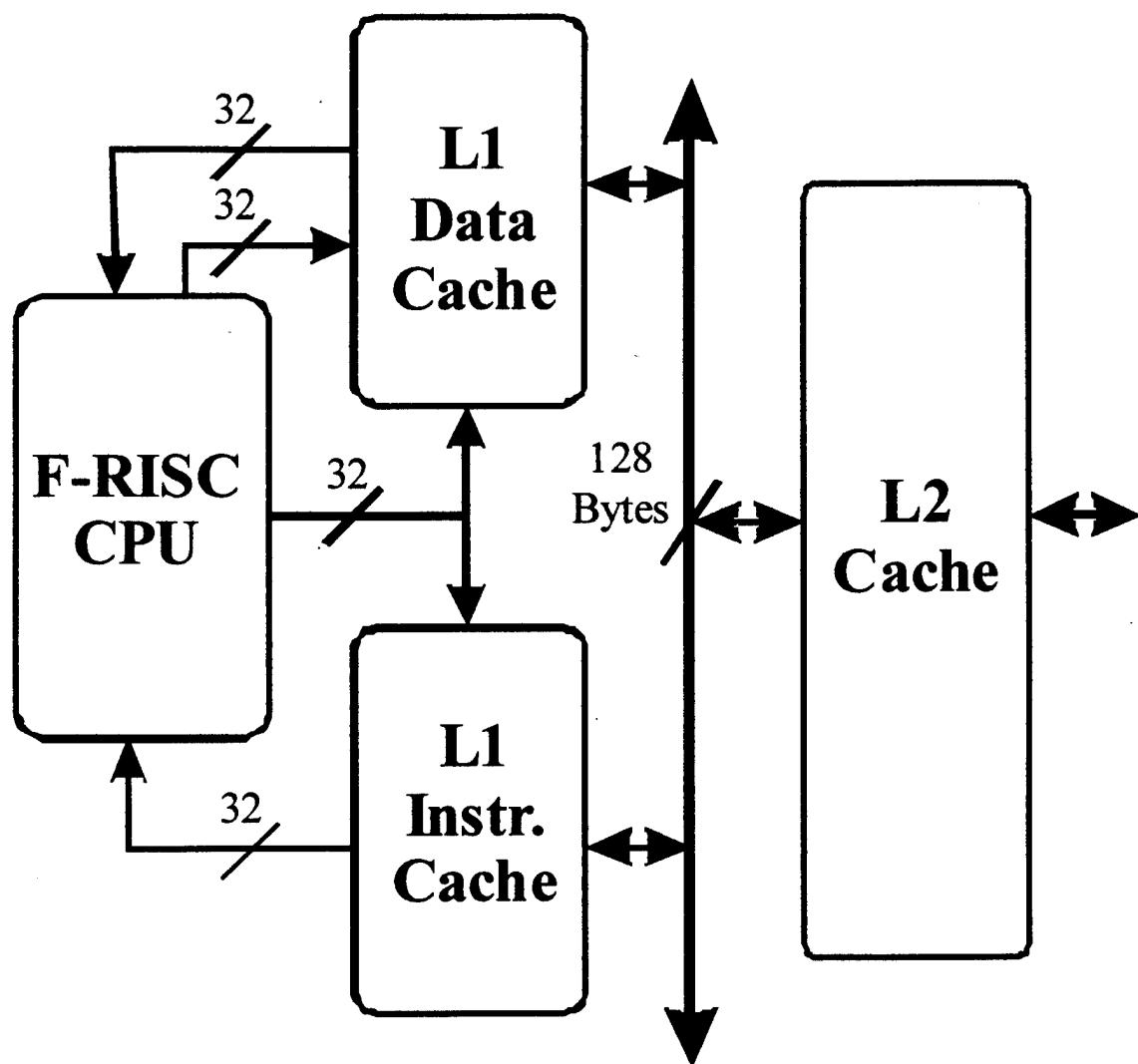
Figure 6. Basic Architecture of the Fast RISC (F-RISC).

Preliminary SPICE models and design manuals provided by Rockwell suggested rather early that a 1 ns machine was possible in this technology. Moreover, much faster HBT's were already being characterized at Rockwell with peak transit time frequencies of 100 GHz, 160 GHz and 320 GHz, and other materials systems such as InP/InGaAs promise even faster HBT's. Hence, as yield and speed evolved in this foundry one could predict

with reasonable certainty that a whole spectrum of subnanosecond computing engines could be developed which would far exceed the capability of CMOS. It is this kind of discovery which the Fast RISC project was initiated to uncover.

These decisions concerning the underlying device and materials systems for Fast RISC research occurred concurrently with decisions in the large mainframe industry to move away from bipolar technology and more towards CMOS. In the short term this trend was justified. However, one may argue that this movement of the industry even further away from the bipolar device and from more advanced material systems contains the possibility that all of the resources of the industry will become totally committed to a single technology that will become increasingly difficult to sustain later, as costs rise, and device or fabrication limits are reached. The cost of fabrication is already too large to sustain a companion bipolar industry, and all research commitments to alternate materials systems have been severely cut back in industry. It is primarily left to university research work to continue to explore alternatives.

II.1.B. The selection of an appropriate architecture for the Fast RISC (F-RISC)

The criteria used for selection of the first F-RISC HBT architecture included yield, heat dissipation, partitionability, and compatibility with known MCM technology at the time of initiation of the project. The initial yield estimates provided by Rockwell to the Rensselaer team suggested that IC's with approximately 5K HBT's could be fabricated with 20% yield. A the time of the initiation of the project there were no IC's of this size with which to confirm that such yield of 20% was actually attainable. The information was gathered by examining clusters of many smaller sized integrated circuits and counting them as a single integrated circuit if there were no faulty components in the cluster. Hence, a key criterion for the architecture, other that it allows fast implementations, is that it also permits partitioning into 5-6K HBT circuits. This restriction forces bitslice or byteslice chip organization and imposes a chip crossing penalty on several critcal delay paths. Additionally the extremely small numbers of transistors per chip forced the design to reexamine many architectural tenets presented by the Berkeley RISC II project. In that earlier project transistors were also available in low numbers which forced a reexamination of every allocation for these transistors. Functions which contributed only slightly to the performance of the system were removed from hardware and shifted to software. Many modern so-called "RISC" processors have moved away from this "guiding RISC principle" as CMOS integration levels have reached many millions of transistors. However, HBT technology also faces this same challenge, plus some more severe ones involving power dissipation. It should be mentioned that during the early phases of this project in 1985, prior to ARPA/ARO funding, Dr. Robert Sherburne, codesigner of the Berkeley RISCII CMOS chip taught for about one year at Rensselaer with the Center for Integrated Electronics after receiving his degree. The influences of this earlier Berkeley RISC II ARPA contract on the present architecture are fairly strong because of this early interaction. The F-RISC project also has a long history, including several earlier academic explorations of embedding RISC processors in other state of the art foundries, including a Tektronix 1.2 μm dual poly bipolar process, with a peak transit frequency of 15 GHz.

Later it was determined that boundary scan techniques would be required to test the chips because of the lack of high pinout probes for testing the completed circuits at speed to identify Known Good Die (KGD) for MCM insertion. This embedded at-speed testing technique would require on-chip circuitry to scan in test patterns at low frequency, spin up one high frequency four phase clock cycle for that test pattern, and scan the results of the

10

test out at low frequency. This meant that only two HF probes would be required for supplying the 2 GHz clock, and another to initiate the four phase cycle. The circuitry to provide this testing capability, particularly for chips with approximately 256 pinouts each, required approximately another 2K HBT's. As the chips finally emerged from various design refinements, their HBT counts had climbed to approximately 8-9K per chip. Fortunately, while the design evolved, the process improved such that these larger chips could be fabricated with yields of 20%, at least if the standard HBT device is used.

II.1.C. F-RISC Architecture

To implement a 1 ns processor with fast, but power and yield limited circuit technologies, a processor architecture is required that can achieve high clock rates, even if the CPU and cache memories need to be partitioned. For example, the 32-bit datapath had to be partitioned into four 8-bit slices that can be implemented with 8-9 k device yields. For the same reason, the cache memories must be implemented with separate cache memory chips. The cache memories need to have a capacity of at least 2 KByte to be effective. In addition, the short cycle time and the MCM delays require subnanosecond cache memory access times. Thus the cache memories must be implemented with the same high speed, but yield limited circuit technology as the processor and hence a large number of cache memories will be needed to implement sufficiently large cache memories.

The processor must be a RISC since RISCs can be implemented with a low device count and support short cycle times through pipelining. A Harvard architecture with separate instruction and data caches is needed to sustain high throughputs by supporting parallel access to instructions and data.

Figure 7 shows different pipeline candidates for F-RISC. A simple 4 stage pipeline IF, DP, D, DW does not allow very high clock rates because instruction decode & operand fetch and instruction execution take place in one DP cycle. The standard 5 stage pipeline IF, DE, EX, D, DW provides a separate stage for instruction decode and operand fetch and thus permits faster clock rates. However, the standard 5 stage instruction pipeline still requires that instruction fetches and data I/O be performed in one IF or D cycle. This constrains the time for an address transfer, cache memory access, and data/instruction transfer to 1 ns requiring a memory access time well below 1 ns. Even with a dense MCM package the address transfer plus data/instruction transfer take a substantial fraction of the cycle time. The delays on the MCM are in the 5-6 ps/mm range, even if low dielectric constant materials are used for the interlayer dielectrics. Thus the transmission line delays alone account for 500 - 600 ps of the cycle time! A 5 stage pipeline would therefore require very fast cache memories which implies low capacity and high power dissipation. However, we can 'hide' the long transmission delays in pipeline stages. The 7 stage F-RISC pipeline provides 2 pipeline stages for instruction and data access allowing a pipelined memory access that allows 500 ps for address transfers and 500 ps for instruction/data transfers. Of course the deeper pipeline also increases the latency of load and branch instructions. The 9 stage pipeline allows 1 full cycle for address and data transfers. Such a deep pipeline will be needed for subnanosecond F-RISC versions.

Instruction Latencies:

Load

Branch

Reg. to Reg.

1 2 2

4-Stage Pipeline

| IF | DP | D | DW |

5-Stage Pipeline

| IF | DE | EX | D | DW |

1 3 3

7-Stage Pipeline

| IF₁ | IF₂ | DE | EX | D₁ | D₂ | DW |

1 4 4

9-Stage Pipeline

| IF₁ | IF₂ | IF₃ | DE | EX | D₁ | D₂ | D₃ | DW | 1 5 5

├────┤ pipeline stage limiting cycle time

➝◄ ►➝ pipelined cache access :
address transfer, cache read cycle, data transfer
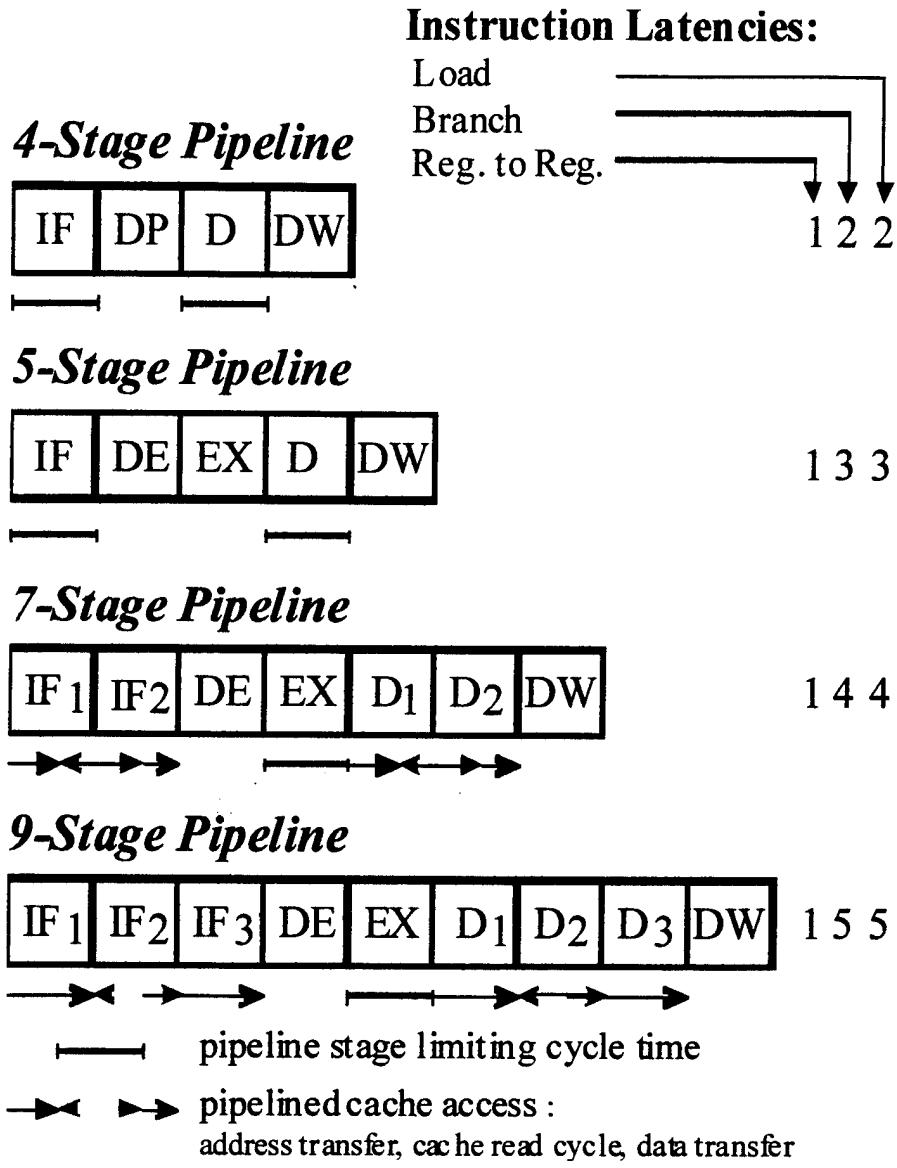
Figure 7. Pipeline Candidates.

The F-RISC instruction set is very regular to speed up instruction decoding and reduce the amount of hardware required for instruction decoding. All instructions are 32 bits long. Instructions with 3 register references (op1, op2, dest) with an optional signed 8 bit immediate constants and 2 register instructions with a signed 16 bit immediate constant are supported. F-RISC has no hardware pipeline interlocks, the full pipeline is exposed. F-RISC provides BRANCH instructions with execute and BRANCH instructions with squash to allow the compiler / code scheduler to reduce the cost of branches.

The main features of F-RISC are summarized below:

- **32 bit RISC**

- **2 GHz clock drives an internal four phase clock generator**

- **highly pipelined to support short cycle times**

- **Harvard architecture with shared address bus**

- **separate instruction and data cache memories**

- **pipelined instruction/data access to 'hide' MCM transfer delays**

- **regular instruction set to speed up decoding**

- **3 register instructions with signed 8 bit immediate constant**

- **2 register instructions with signed 16 bit immediate constant**

To obtain the extreme speed required to keep feeding instructions to the processor, HBT technology had to be selected for the first level (L1) of the cache memory. This immediately implied a small off chip cache memory. To avoid the huge penalty resulting for a high miss rate in L1, the penalty for a miss was reduced dramatically by making the transfer of data or instructions from L1 to the second level (L2) of memory more efficient (meaning wide). A path was created that was 1024 bits in width between L1 and L2, making it possible to transfer an entire cache block in one L2 memory cycle. Differential I/O with power balanced open collector drivers is employed to reduce switching noise and reduce driver delays.
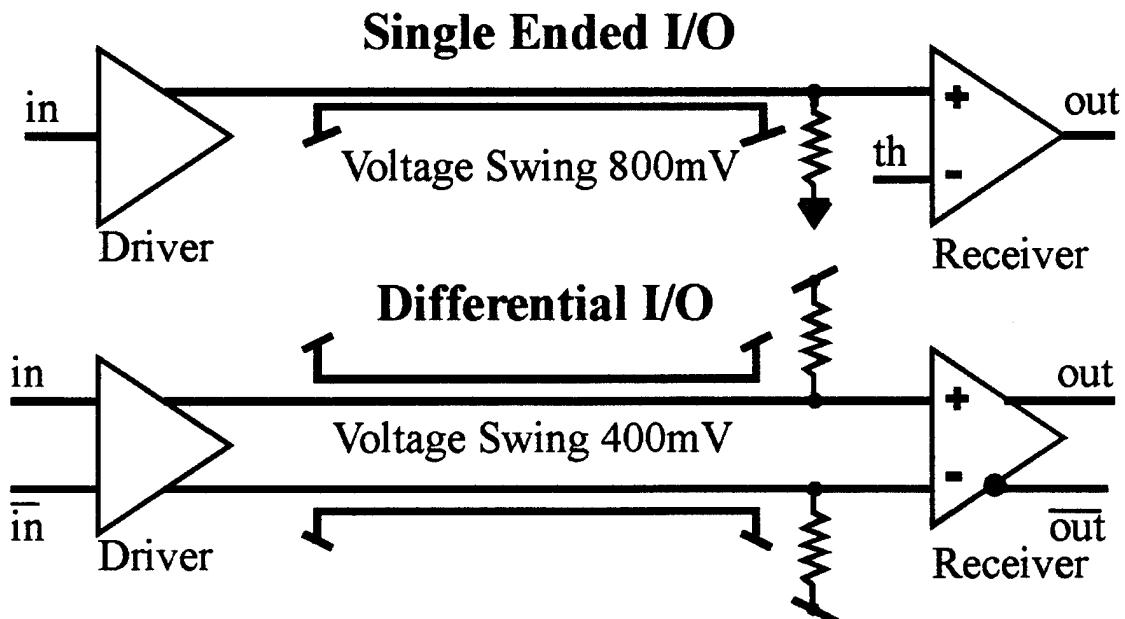

Figure 8. Differential I/O Circuits.

13
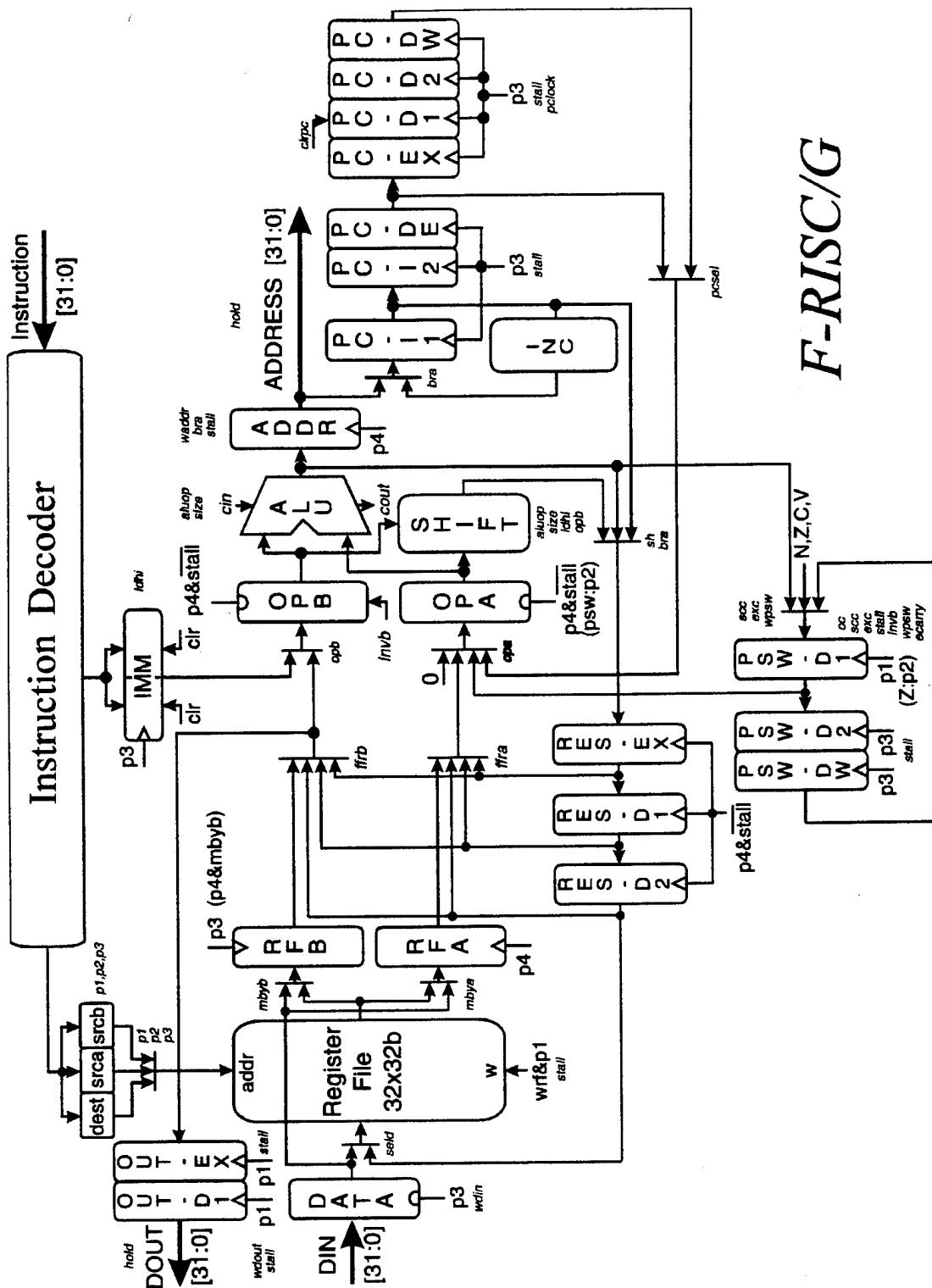
Figure 9. Internal Pipelining, data path, and component structure of the F-RISC Architecture.

14

II.2. SUMMARY OF THE MOST IMPORTANT RESULTS

The summary of research activity during the first three contract years followed roughly the plan presented in the contract proposal:

• In the _first year_ a standard cell macro library was developed using design rules and models provided by Rockwell. Over one hundred twenty cells were developed and tested for the library. In addition several large memory block macrocells were developed. Computer Aided Design (CAD) tools were developed to facilitate the design of full differential Current Mode Logic (CML) circuits with closely tracking wire pairs. Full differential CML offered significant capability to eliminate the switching noise associated with single ended logic, and permitted differential suppression of EMI and coupling, which are important at HF.

• In the _second year_ of the project the 5 GHz 8 bit x 32 word register file (RF) for the Data Path (DP) chip was designed. This component of the design contained some of the fastest signal paths in the architecture and was extremely sensitive to wiring capacitance. Consequently it was designed as a hand crafted large macro. At Rockwell's suggestion a partial reticle test chip design was undertaken to attempt to probe the yield and speed of the architecture. Designs for the Data Path, Instruction Decoder, Level One Cache, and Cache Controller chips were begun. Chips of this complexity take about two man years to complete each. Extensive simulations are required to establish that the chips are designed to be functionally correct and that they will work at the desired speed. Functional correctness was guaranteed by multiple chip FPGA emulation using APTIX programmable circuit board technology.

• During the _third year_ the test chip, which was the first fabricated by the group, was returned to Rensselaer for testing. The test chip was created to write random patterns into random addresses of the register file, and reread these subsequently to verify that the write and read produce correct results. In the same year work on the DP and ID chips were completed and a Phase Locked Loop (PLL) clock deskew chip was completed. The clock deskew scheme is critical to guarantee synchronous arrival times of all clock edges at all chips regardless of their position in the ultimate MCM. In addition, two chips were designed for the cache and cache controller chips for the level 1 instruction and data cache memories.

The result of this work is shown in Figure 10 which shows the four architecture chips assembled into a reticle for fabrication at Rockwell. The following figures show the architecture of the F-RISC testchip, a micograph of the test chip, the microwave test setup, our Tektronix probestation with CASCADE 5 GHz six channel probes, and an LFSR waveform at 2.3 GHz and a memory test waveform at 1.2 GHz.

15

F-RISC Architecture Reticle

Datapath.
Chip

PCM

Instruction
Decoder

Cache
Chip
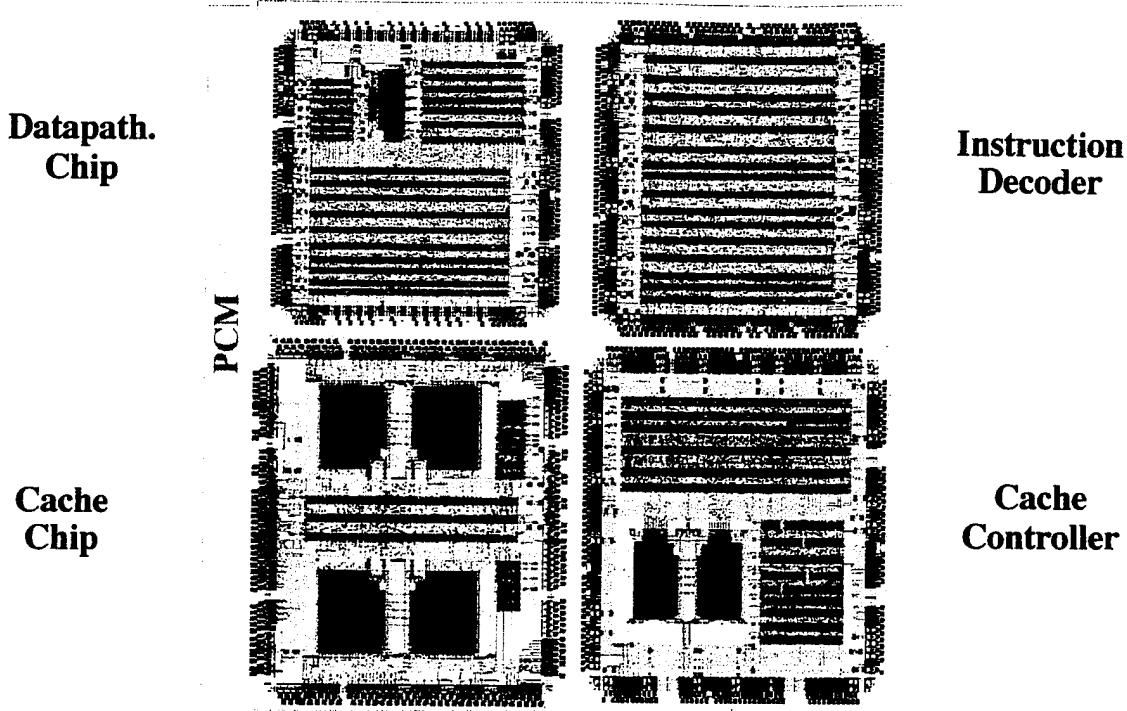
Cache
Controller

Figure 10. Reticle overview containing artwork for the four byteslice architecture chips of the F-RISC: The Data Path chip (DP), The Instruction Decoder (ID), the L1 Cache Memory (CM), and L1 Cache Controller (CC).

II.2.A. RPI Test Chip

Encouraging results were obtained on the test chip in the sense that all subcircuits in that system were found to work, notably linear feedback shift registers, address decoders, registers, multiplexers. and adders. These results validated the cell library and the earlier work on the CAD tools for differential routing and wiring. However, the yield was disappointing, with typical circuit sizes of only 300 HBT's, considerably smaller than expected. Rockwell personnel indicated that this would be greatly improved as they upgraded the Newbury Park fabrication line to 4 inch wafers and introduced a brand new I-line stepper. It was assumed that this yield problem was anomalous. Nevertheless, another disturbing result was that all circuit speeds were slower than expected based on the Rockwell supplied HBT SPICE model and design rules given in their design manual.

16

F-RISC/G Test Chip
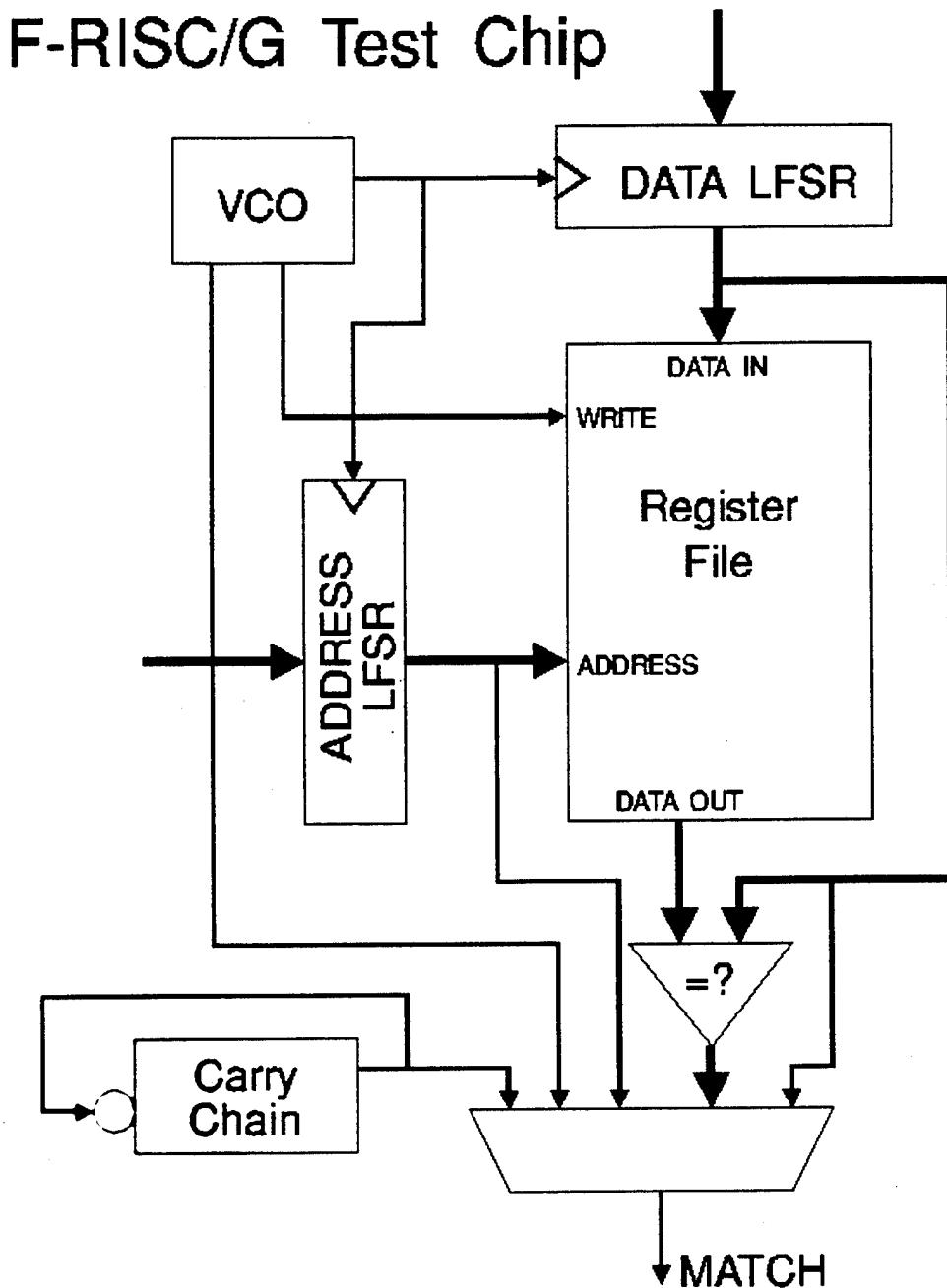
Figure 11. Architecture of the first RPI Test Chip, showing a Register File (RF) and associated test circuit with LFSR address and data generators. Based on SPICE simulations with the Rockwell supplied HBT model this circuit should have run at 3 GHz.
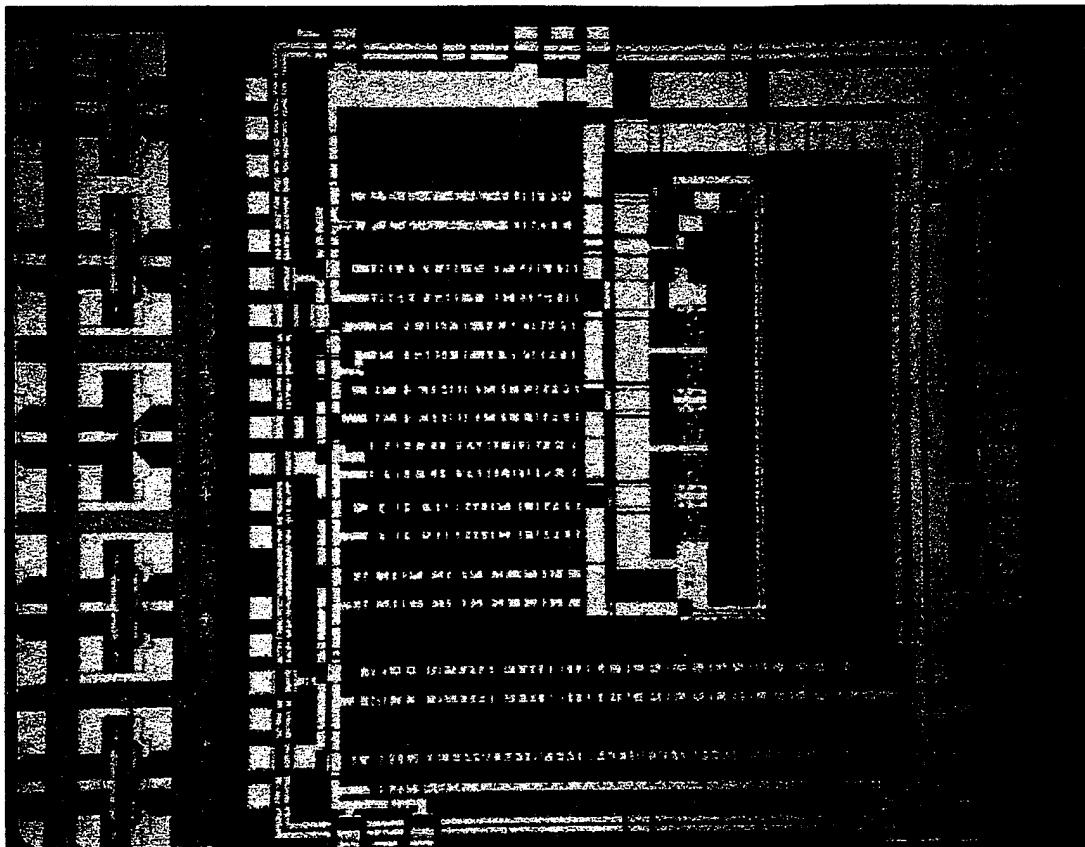
Figure 12. The Rensselaer "Test Chip" as fabricated at Rockwell. Right upper dark region is a dense hand crafted 32 w x 8 bit Register File, while the circuitry to the left contains two LFSR and VCO circuits implemented with standard cells.

These speed degradations ranged from 33% in lightly loaded subcircuits to nearly 50% in circuits with more significant capacitive wire loading. This speed deficiency meant we could not commit our major architecture foundry funds until the anomaly could be explained and a strategy devised to recoup the speed. It was felt that a 500-660 peak MIPS (1.2 GHz clock) F-RISC would not demonstrate a performance range of computers faster than CMOS could attain. Hence, unless this speed problem could be addressed, the F-RISC project would not break sufficiently new ground. This speed problem prompted a request for several no cost extensions to preserve the foundry fee to fabricate the reticle shown in Figure 3 until a satisfactory solution could be found that would guarantee the speed result that was expected.

18
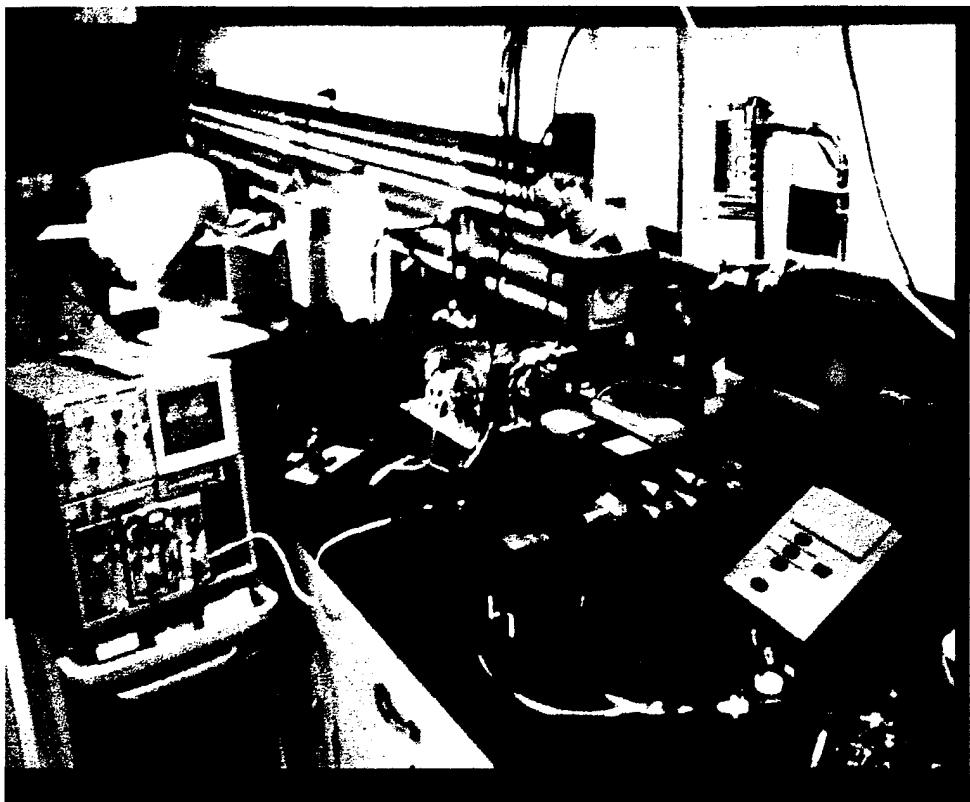
Figure 13. Test setup for at-speed test of F-RISC test chips.

Figure 14. Close up view of CASCADE and GGB probes.

19
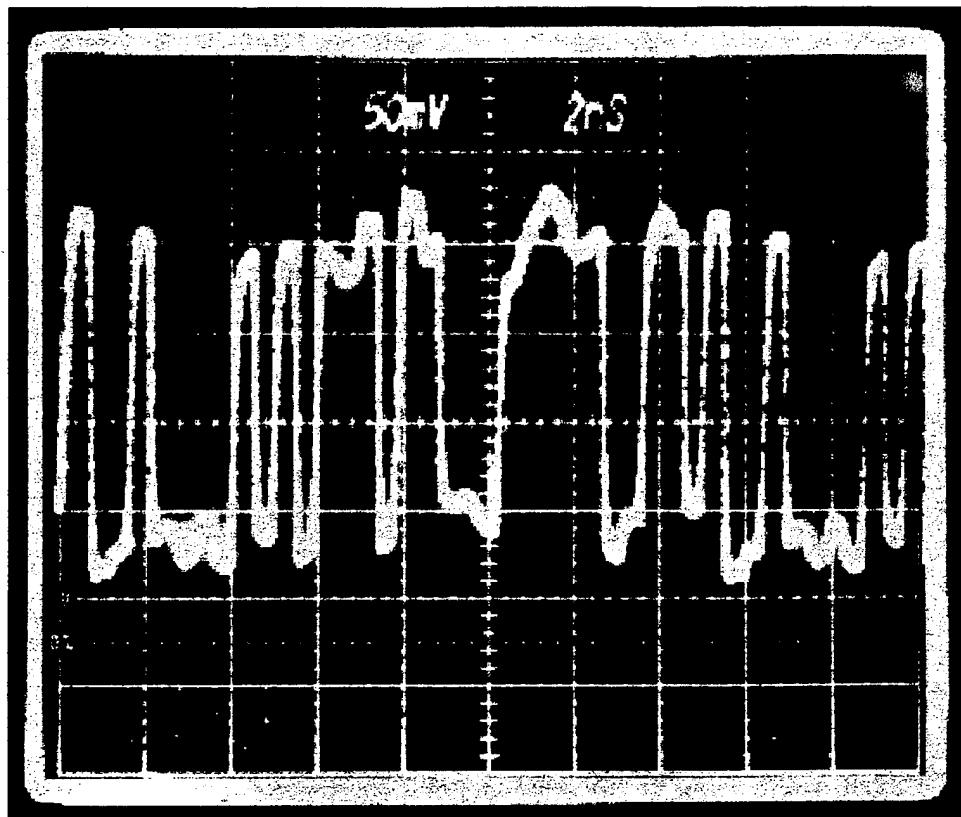
Figure 15. Fastest observed LFSR waveform from the first RPI test chip fabricated at Rockwell. Clock rate is 2.3 GHz, or about 33% less than predicted.

An early S-parameter set provided by Rockwell for an isolated transistor in the PCM for our first test chip wafer lot indicated that the HBT's exhibited about 33% less transit time frequency than expected. In addition, a prescaler circuit on the same fabrication run for another user ran only at 11 GHz rather than the 16 GHz expected, also exhibiting a 33% degradation. Our contacts at Rockwell thought this to be an aberration, and not a cause for alarm. This still left the unexplained wiring delays to analyze for heavily loaded circuits. Sections of the Fast RISC architecture are extremely heavily wire loaded, especially the register file (RF) which has long vertical and horizontal bit and word lines that exhibit large capacitance values dictating the speed of that critical component. The 5 GHz register file, or at least the columns that were testable (with a 2.5 GHz designed clock using 200 picoseconds of up and down going clock phases) was operating at the 50% degraded speed of only 1.2 GHz.

20
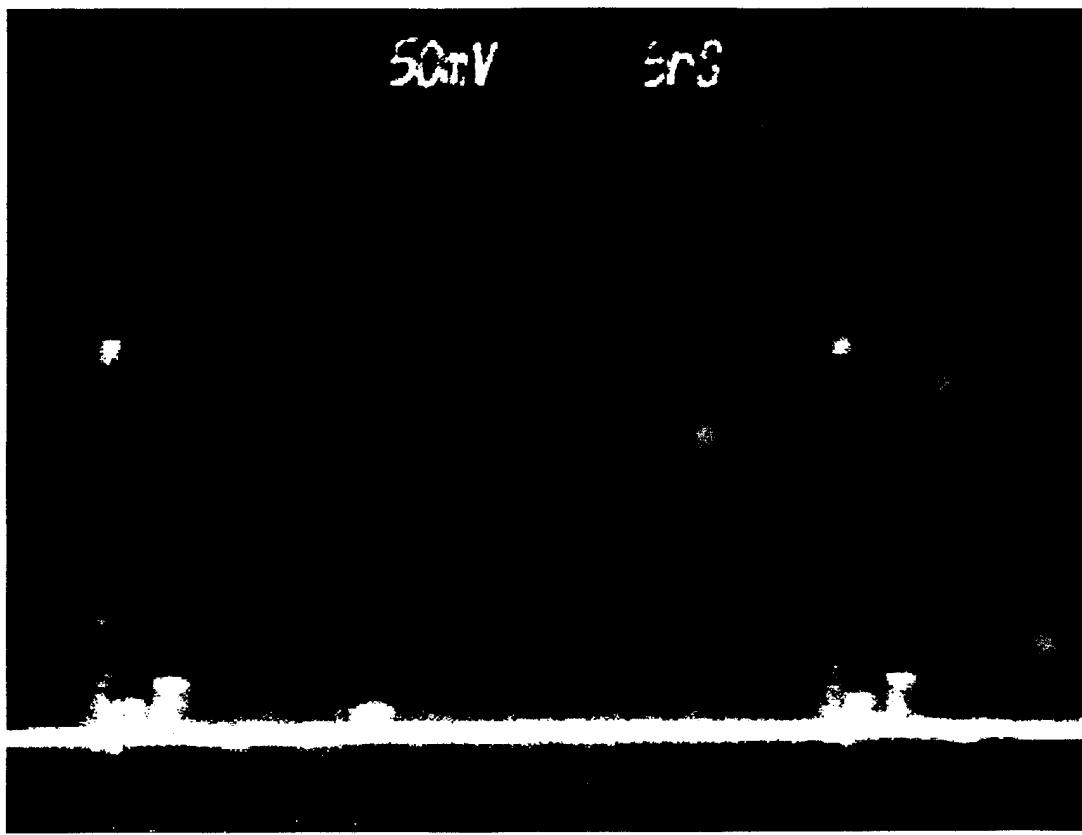
Figure 16. Error Comparator Readout for Register File just at the frequency when the first error begins to appear at a frequency of 1.2 GHz, translating to an access time of 400 picoseconds.

This brought a more critical review of the register file macro. It required a completely redesigned core memory array to reduce the anticipated worsening of bit and word line capacitances. Additionally the design of the Cache memory chips had been contingent on using this same register file macro. However, the redesigned register file grew hotter with each iteration in the design process. Furthermore, the testing scheme chosen for the chips for selection of Known Good Die (KGD) for MCM insertion was a variation on the scheme known as Boundary Scan to test the chips at speed, or At-Speed Boundary Scan (ASBS). Test patterns could be scanned into the chips, intercepting the chip pad input paths, at slow speed, and upon completion of this scan, the chip would "spin-up" for one or two 4 phase clock cycles using the 2 GHz clock using a small state engine. After this the result could be scanned back out of the chip along the pad boundaries. This circuitry adds a burden of approximately 2,000 HBT's to each of the four "byte-slice" architecture chips. Since at the end of the first three years the cache memory chip was emerging as the largest chip with nearly 10,000 HBT's including ASBS circuitry, it clearly would have required use of redundant memory blocks to yield, since the expectation was for 5,000 HBT circuits to yield at 20%. Additionally the heat for this chip (many of them were required for the architecture) became excessive. Introduction of redundant register file blocks and associated multiplexer selection circuits would clearly drive the power dissipated into an unacceptable regime. The indicated solution was to depart from using the "safe" register file

21

from the Data Path chip in the Cache as a macro, and to develop a more power efl design.

At this point in time contract funds for salaries were nearly expended. Follo ARPA/ARO contract work, a companion AASERT contract, and an HSCD Roc subcontract helped provide the manpower to redesign the register file and cache m block. However, foundry fees for the fabrication were preserved through a series (cost extensions, while work on the circuit revisions proceeded. Important addi support came when Rockwell selected Rensselaer to participate in its HSCD BA/ design group and cell library development group. This helped provide access to addi partial reticle fabrication runs, and brought more manpower to the group to pursu what the exact nature of the speed problems were in the Rockwell process.

II.2.B. HBT Device Models & Switching Performance

A device modeling problem was detected in the Rockwell process througl participation in the HSCD project. Unloaded ring oscillators on the first HSCD run found to run slow by 33%. Hence the HBT itself exhibited a problem, exclusive ‹ previously discussed ILD thickness control problem. This took the greatest amount o to investigate because initially such problems were not expected. Hence, all the early test circuits did not contain test structures to probe and model the HBT. The I funding provided a mechanism to explore this problem in some detail. But th‹ indicators of a problem were found on the first RPI test chip fabricated in year three that reticle Rockwell was able to give us an S-parameter measurement of the HB program was developed to "fit" SPICE parameters to this data by using SPICE to sii the generation of the S-parameter data, whereupon a direct comparison could be m the measurements. Even though the bias point on the collector emitter voltage Rockwell S-parameter measurement was not ideal for our circuit's range of opera could be determined that the transistor "behaved" as if it had a 33% lower transi frequency at all collector current values less than the dopant redistribution limit f transistor. Since the plot of this frequency for various collector currents is inv proportional to the total base capacitance of the HBT, this implied that the C_b wa: larger than the SPICE model provided in the Rockwell design manual, and that mo this had been the case for all the years into the design cycle.

Figure 17 and Figure 18 compare the magnitude of S21-Parameter measurements on devices from the first HSCD run with S-Parmeters of different device models. The S-parameter measurements have been made by Mayo on an RPI test structure. The measured S21 parameters which are an indicator of the gain and bandwidth of the device are compared with the S21 of the device model in the design manual (S21_q1_dm) , ˆ ˉˉˉ switching device model and a 33 GHz model extracted from the RPI testchip fabrica (S1_q1_33). The 33 GHz caused initial speculations that the process was off again s predicted circuit speeds much more accurately than the official model in the desig manual. The switching models have been recently developed by Rockwell in respc RPI's closed loop design & simulation and testing work that proved that the model design manual was off by 33% in predicting switching speeds.

22
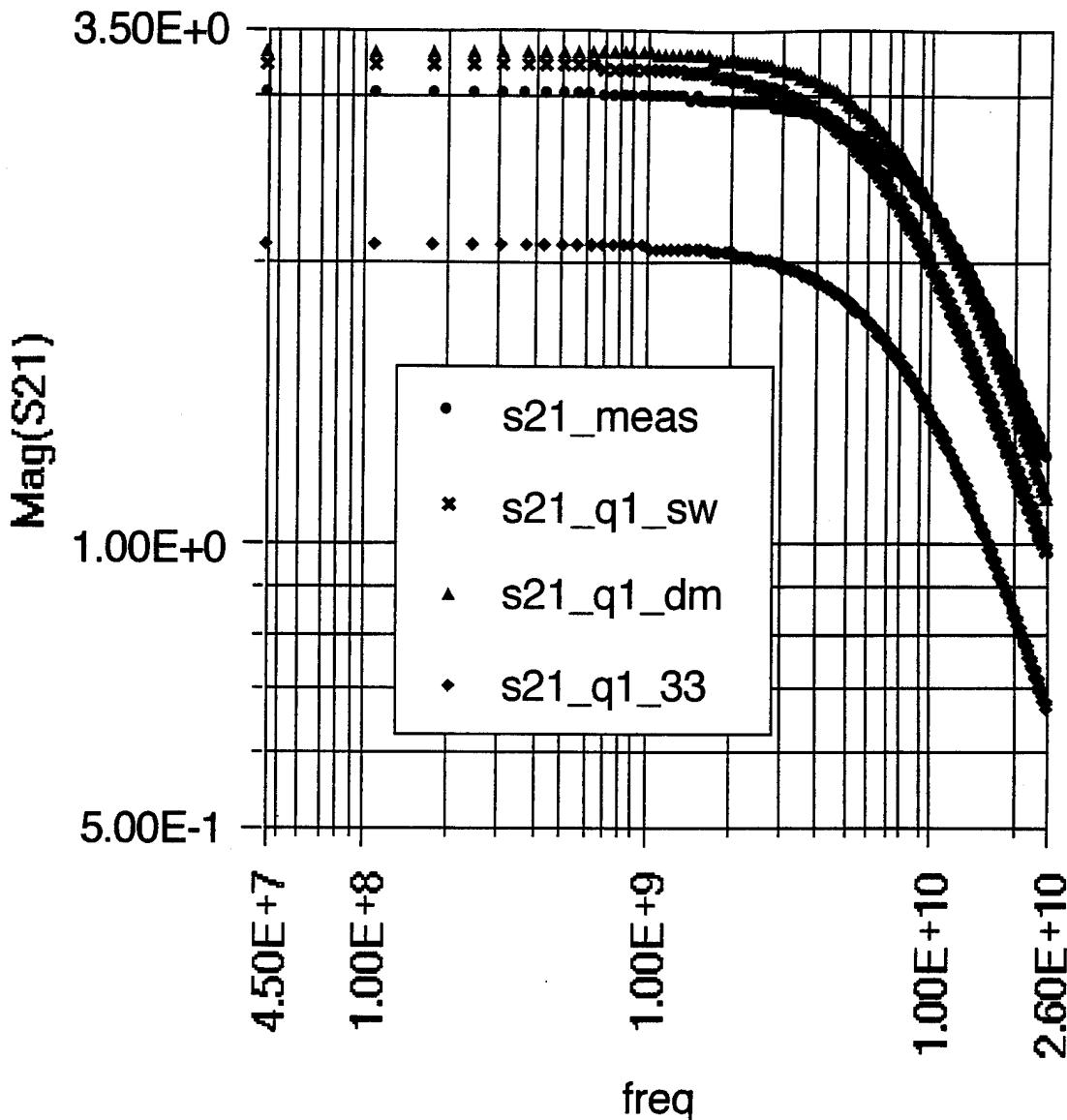
Figure 17. Measured and Model S21 Parameters Compared (Ic0=2.1mA, VCE0=2V).

The measured S-parameter match the model quite well at currents levels (2.1 mA) were the device reaches optimal F_t . However, in switching applications the devices are turned on and off. Hence, not max F_t is relevant, but how quickly the device turns on or off. The turn-on characteristics of the device are most important for the switching time since the device spends most of the switching transient in the low current regime since the device is much slower at low current than at high current levels. This correspond to the F_t or S21 parameters at low current levels. The following figure shows that the measured S-parameters on the first HSCD reticle run (Dec 94) are still much lower than predicted by any model at low current levels (0.4 mA) . Part of the problem with the SPICE models is that the Gummel-Poon SPICE model is not an good fit for HBT devices. The new SPICE model under development under the HSCD program can match the measured characteristics much better both in the high and low current regime.
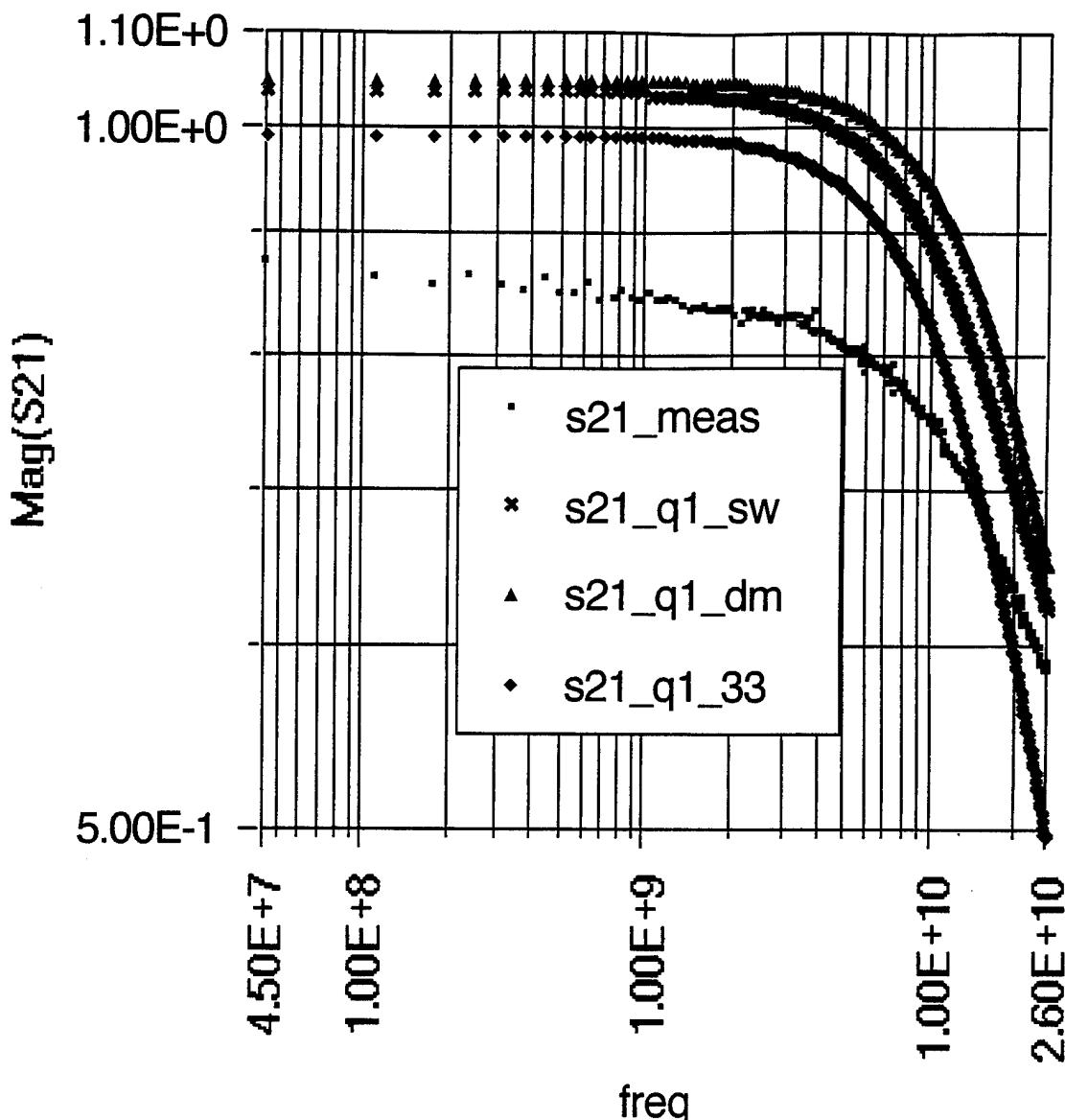
Figure 18. Comparison of Measured and Model S21 Parameters at Ic0=0.4mA VCE0=2.0V.

II.2.C. Interconnect Capacitances and Interlayer Dielectric Thickness

In the *fourth year* following the initiation of the subject contract a special 3D capacitance extraction program was developed. The program, an outgrowth of another Professor's work at Rensselaer is termed QuickCAP. Professor Y. LeCoz is its developer. This program was found to be the only program available to the group which could perform detailed 3D capacitance extraction for conductors in wiring channels or macrocells such as the register file. Entire wiring channels could have *all* their conductors analyzed for the complete capacitance matrix in a format suitable for use in SPICE. Using this tool the DP register file was completely redesigned for its intended 5 GHz operation. A new third level of metallization was incorporated into the design. In addition a 10% slack was incorporated into all timing to enhance the chances for success of the project. Furthermore

24

the core memory block (MB) in the cache memory was completely redesigned around a 16 bit x 32 word organization to reduce the number of HBT's required in address decoding that were employed in the DP register file. The errors detected in the original design included computed capacitance values that were off by 200% in some cases due to 3D effects. It was expected that this might explain some of the circuit speed degradation in heavily loaded circuits. Reduction in the number of HBT's helped reduce power and increase the yield of the cache chips and their controller.
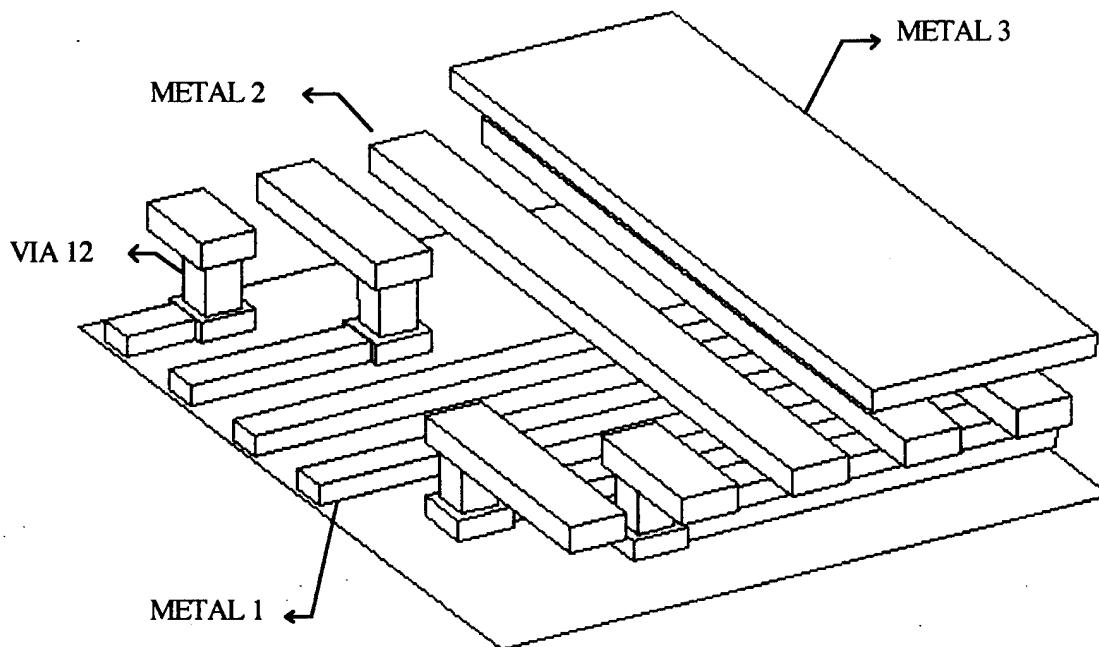
Figure 19. Sample 3D interconnection structure in the vicinity of a standard cell routing area and a power rail crossing illustrating several complex geometric effects that must be included in capacitance extraction programs to get accurate circuit delays.

Concurrently an effort was launched to create a variety of test structures which could be employed to verify that the newly recalculated values of capacitance were correct. Numerous ring oscillators were constructed under HSCD funding and submitted under a shared reticle fabrication run to probe the speed of these circuits. Some ring oscillators were unloaded while others were loaded with a variety of capacitive wiring structures. These were fabricated toward the end of the fourth year and tested extensively at RPI, the ARPA high speed group at the Mayo Clinic, and Rockwell. Among these structures were several large area capacitor structures created between different levels of the metallization layers (now three in number).

The first stunningly simple result was that these large area capacitors, created simply as an afterthought to check dielectric thickness, showed anomalously high capacitance by factors of from 45% on M1-M2 layers to 54% on M2-M3 layers. The capacitors were actually large enough to use the simplest formula for computing capacitance with less than

0.5% error. Since the M1-M2 capacitance was 45% high it suggested that the dielectric thickness or dielectric constants were off. Since conventional DuPont 2610 polyimide had been used as the M1-M2 interlayer dielectric or ILD, this suggested a dielectric thickness of only about 70% of the design manual value. Rockwell's published nominal thickness was 1.6 microns for this layer of the ILD. The measured capacitance values suggested that the thickness for large area capacitors (about 200 microns by 100 microns) was only 1.2-1.3 microns thick. Rockwell's standard fabrication calibration is to check this thickness at 5 scribe lane locations. Rockwell pursued this further and found that at certain locations inside our dies the ILD thickness between M1 and M2 at a standard width wire crossover was only 0.9 μm. This variation of nearly 50% in thickness was much larger than expected.

However, due to the differential wiring scheme used in the F-RISC circuits, and due to the semi-insulating substrate, most coupling field lines are horizontal between wire pairs. This can be seen in the following figure wherein it is shown that a great number of the field lines are approximately horizontal.
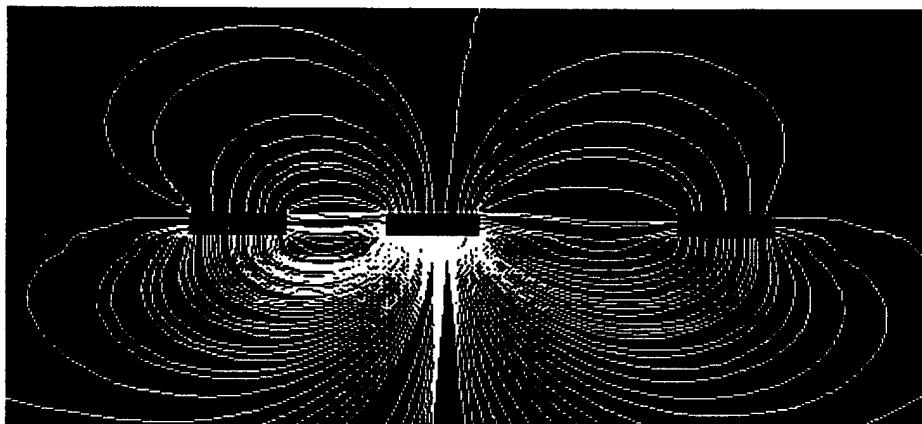
Figure 20. Electrical Field analysis for parallel conductor assembly of three interconnections over a GaAs substrate.

Therefore the impact of the greatly thinned ILD is less than one would first think. Hence even such a large deviation in thickness from the nominal value could produce only about a 15% increase in wire capacitance if this alone were the problem. Unfortunately polyimide is an anisotropic dielectric with about a 10-15% higher dielectric constant in the horizontal direction due to the fact that polyimide is a polar material and the polymer strands lie horizontally in the film. Consequently, the combined effect of both the thinner ILD and the anisotropic dielectric constant could produce net excess capacitance in differential wire pairs by 20 to 25%. Rockwell advised that it would not be able to alter this situation quickly, and so a strategy had to be devised to offset this deficiency.

Fortunately, the delay in fabrication of the architecture chips had permitted Rockwell time, however, to make several other process improvements which are early introductions of some aspects of their proposed 100 GHz process. One of these is a shrink of M1 metal wire widths from 2.4 to 1.6 μm. This shrink was accompanied by a reduction in wire separation rules also, which would permit reduced wiring pitch and wiring length. However, to offset the increased capacitance due to the aforementioned

thickness variations and anisotropy it was shown that decreasing wire width to the new rule, but not adopting the new wire separation rule would fix the excess capacitance problem. This approach would leave the wiring pitch the same, while decreasing the horizontal field component of the wiring capacitance by enough to essentially neutralize the increases. Additionally, some M2 power busses could be removed from the macrocells leaving only the M3 power straps, considerably increasing the distance between M1 and any top metal ground plane. Since it is expected that Rockwell will eventually fix the ILD thickness uniformity problem, and perhaps introduce more it is felt that these two changes in wiring capacitance provided a reasonable compromise interim measure. In the course of making these alterations, it was discovered that narrowing some of the longer lines in the architecture started to make the self resistance of these lines more noticible. Some of these have had to be relocated manually to the M3 level where metal thickness and dielectric thicknesses are about three times larger than for M1.

II.2.D. New Switching Devices with Lower Junction Parasitics

Test circuits developed on early HSCD funding helped confirm and refine the RPI version of the SPICE model for the 1.4 μm x 3 μm emitter stripe baseline HBT, which also found differences in other SPICE parameters. However it was not until the *fifth year* of the contract that enough information had been gathered to address possible HBT changes with any confidence. The model discrepancy discovered in this manner showed that the base capacitance is extremely important during the turn-on phase of the HBT when the collector current is low. Since the CML circuits must switch the transistor from zero current to some nominal value, the behavior of the circuit for low collector current tends to dominate the switching time. The 33% larger base capacitance is observed only in this turn on regime. Apparently this discrepancy was not known by Rockwell during the development of the model, which had its origins in analog circuit designs where collector bias currents are typically set to get optimal device performance. The F-RISC project was more sensitive to this problem than other circuit designs since the project had a specific speed goal. Rockwell has been extremely helpful in every way possible to accommodate the requirements of our project in view of this model deficiency including providing information on some aggressive transistor layouts they had considered.

One limitation of this device research has been that no process alterations could occur (no doping levels, thicknesses, or alloy ratios could be changed). Therefore any solutions possible had to be effected through the layout of the transistor. Since layer compositions and thicknesses for the epitaxial layers were not revealed, these alteration steps had to be estimated. Device modeling programs such as TMA, Inc. DAVINCI or SILVACO UTMOST are of only limited use without disclosure of these parameters. Nevertheless, work is in progress on using these programs to gain insight about trends likely to be seen when varying various parameters.

New Shrunk Wiring **Old 50 GHz Process Wiring**
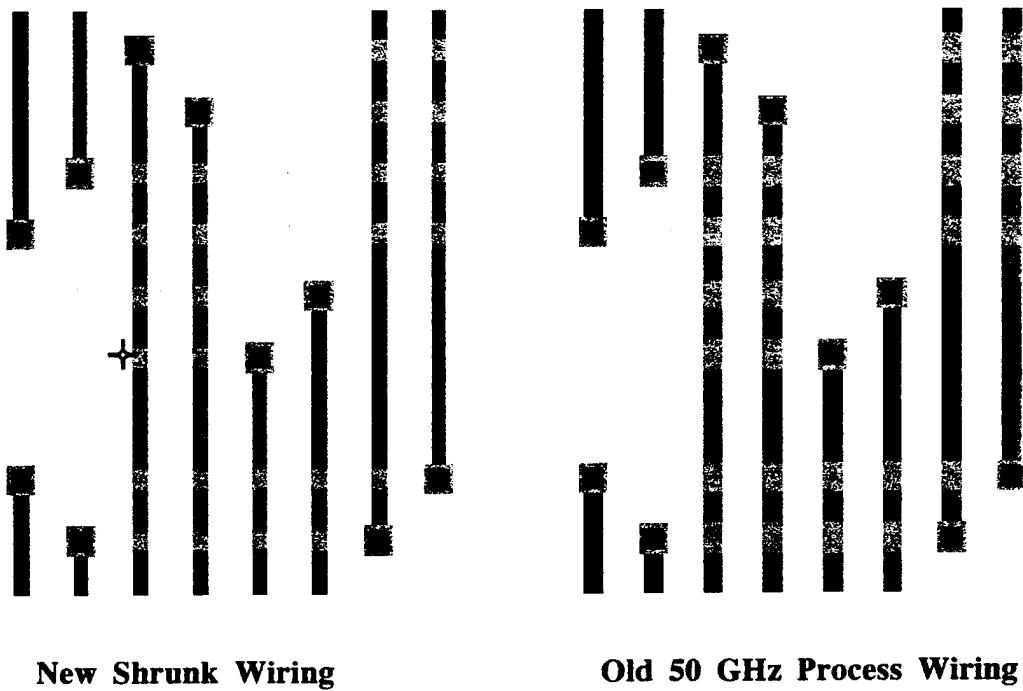
Figure 21. Partial use of the narrow wire width design rules (left) of the Rockwell 100 GHz process keeping the same wiring pitch of the 50 GHz process (right).

The primary parameters to which designers have access is the layout of the features of the transistor, such as the emitter stripe area, base to emitter separation, base pedestal area, base contact area, and location of the collector contact, moat and collector definition.

Of all the accessible layout features such as the emitter stripe area, and the base pedestal area have the largest impact, because SPICE simulations show that the base capacitance is the leading parameter affecting speed. However, base resistance and emitter resistance can impact the amount of current going into the base, and hence through the collector. Since the designs are completed and only the transistor layout can be varied without performing large amounts of redesign, which would require several man years of effort. We note, however, that a fresh design project would not suffer from this carry over, and a larger base resistance could be by designing the circuits for a slightly higher base voltage swing.

The following figures show the standard HBT device with an emitter size of 1.4 μm x 3 μm and several RPI device layouts with an emitter size of 1.2 μm x 1.7 μm. Test structures with these devices are or will be fabricated to evaluate performance and yield of these devices. Rockwell is persuing the round emitter device shown in Figure 23 under the HSCD program. However, ringoscillators on the RPI testchip did not indicate that the round emitter devices provide faster switching speeds.

q4p20p

q2p04a
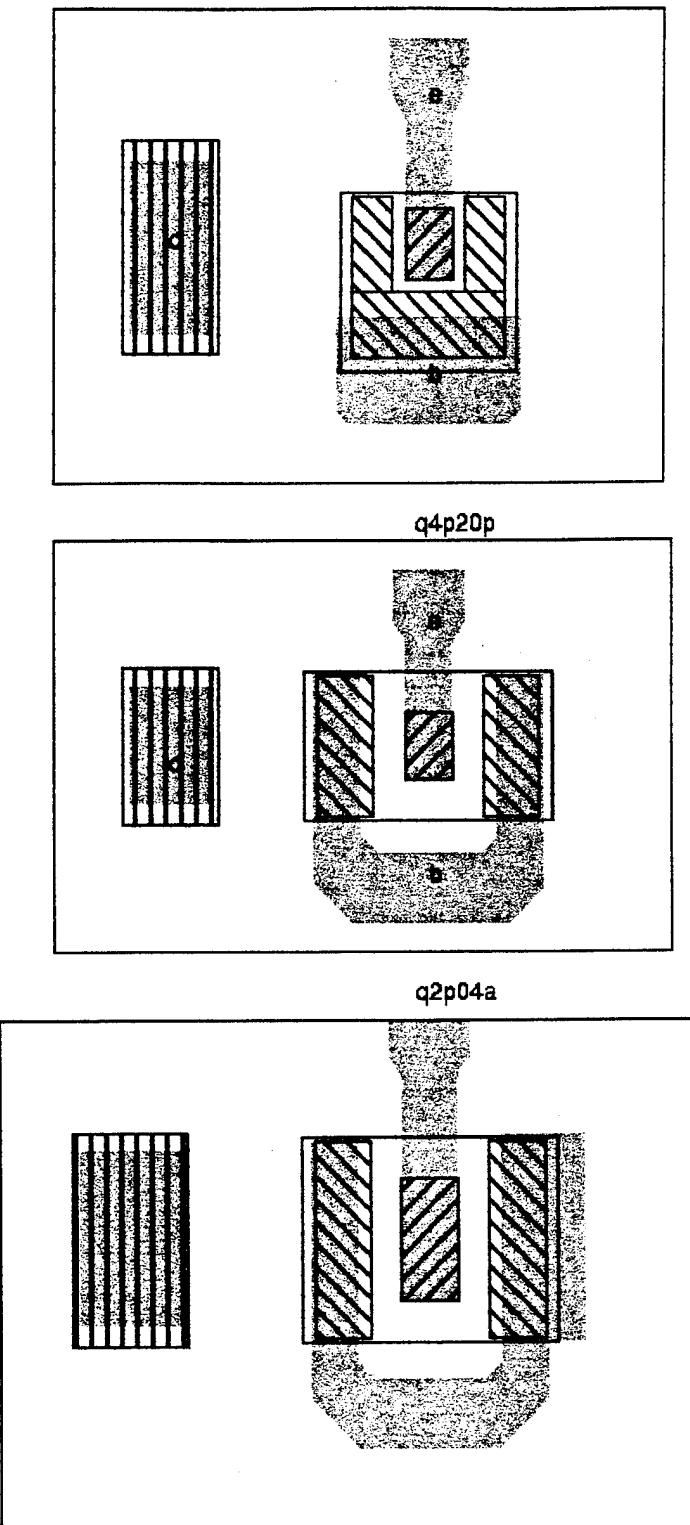
Figure 22. Evolution of the "50 GHz" basic HBT. Lower HBT is the "original" 1.4 μm by 3 μm emitter HBT supplied by Rockwell in its design manual. The middle transistor has the emitter shrunk to 1.2 μm by 1.7 μm and shortened collector base separation. The top transistor is an aggressively scaled device layout with a 1.2 μm by 1.7 μm emitter and a 0.4 μm base-emitter separation.

29
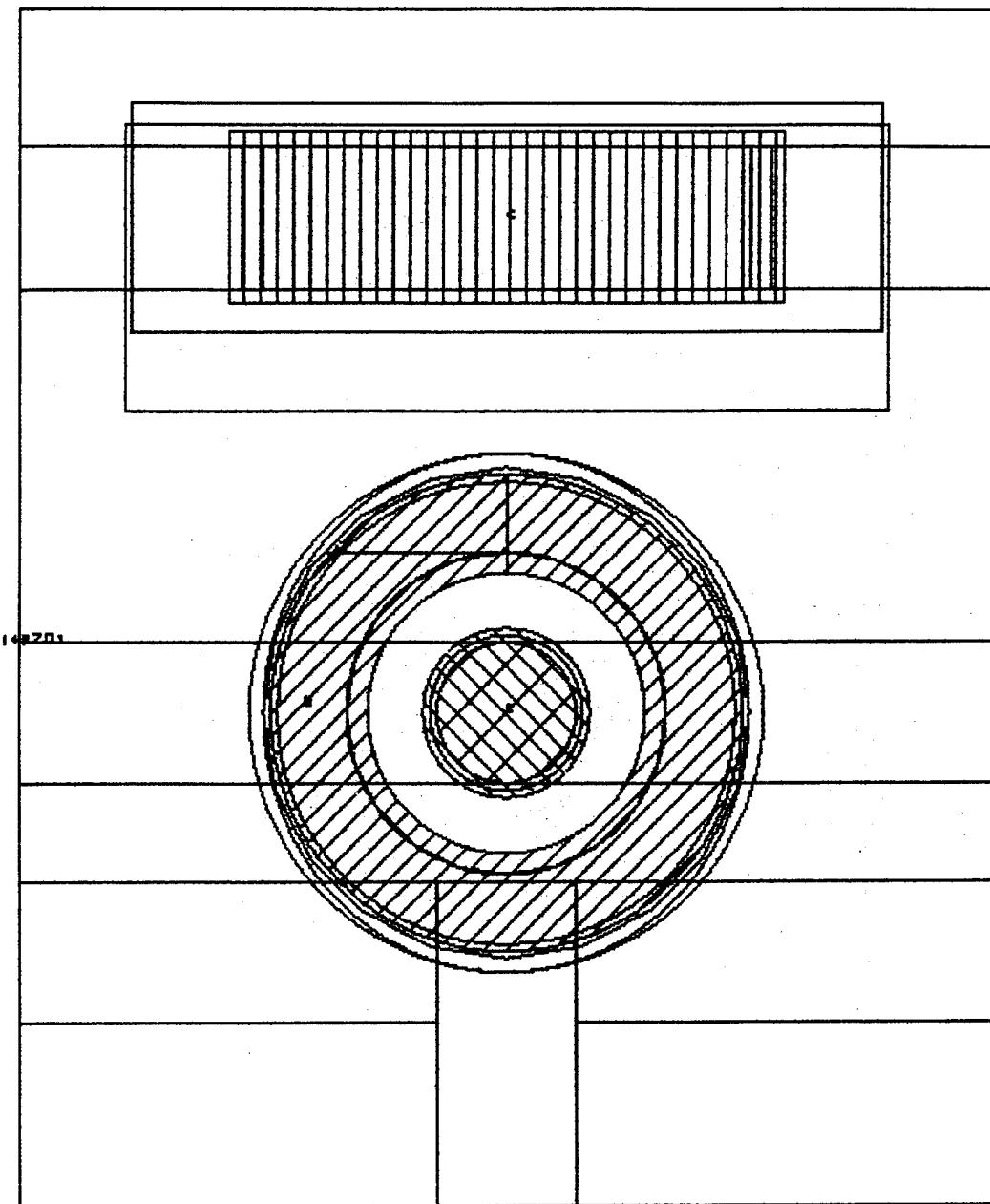
Figure 23. Q4P20FA Device with Round Emitter (D=2.3μm).
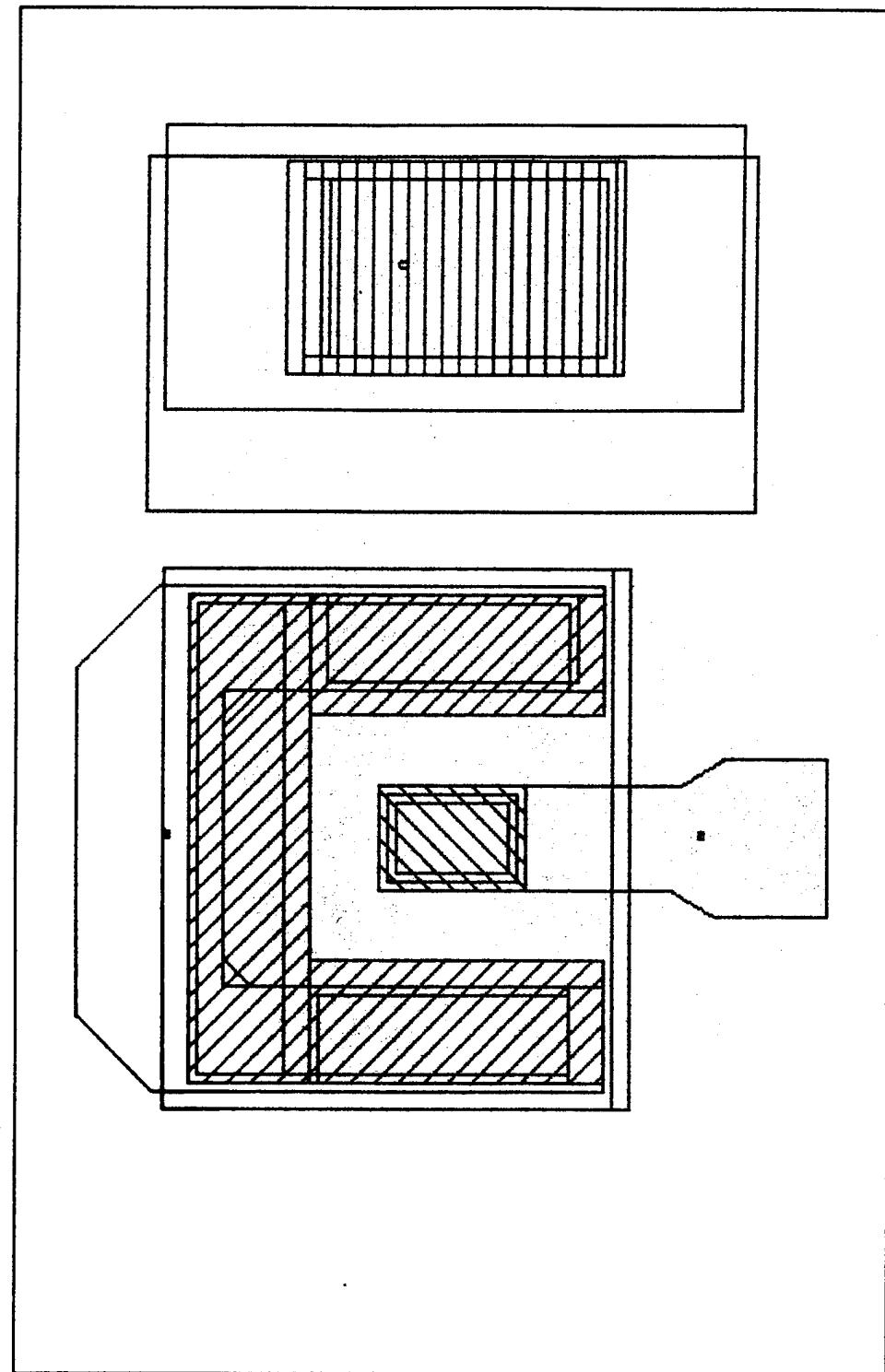
Figure 24. Q2P04 Device with Base Contact on Third Side and 0.8 μm minimal Spacing, Scaled Emitter = 1.2 μm x 1.7 μm.

When the emitter stripe is shrunk the component of the base capacitance resulting from the base emitter junction will decrease proportional to the shrinkage of the area of the emitter. But the base and emitter resistance then increase. The emitter resistance arguably increases inversely with the area shrinkage because the current flows vertically through the emitter. This is how the SPICE "AREA" parameter changes both the base and emitter resistance when the emitter area shrinks. However, fortunately the emitter resistance is small compared to the base resistance even with such a shrinkage.

For the base resistance, the intrinsic portion roughly grows with the shrinkage of the emitter area, and the extrinsic component grows inversely with the perimeter of the emitter area all else remaining the same. Unfortunately, the exact partition of the base resistance into its extrinsic and intrinsic component are difficult to predict without detailed layer information. Rockwell estimated this ratio of extrinsic to intrinsic base resistance to be 4:1, illustrating the importance of the extrinsic portion. Hence as the emitter area is shrunk one would like to maintain the perimeter of that area. Rockwell assisted us in the evaluation of a series of potential substitutes for the original 1.4 m µm by 3 µm emitter stripe (4.2 square µm area) HBT offered in their baseline process. From this collaboration the first evolved HBT was developed reduced the area of the emitter stripe from 1.4 µm by 3 µm to 1.2 µm by 1.7 µm (2.04 square µm area or approximately half the area of the 50 GHz baseline HBT).

This emitter scaling was only possible because of a switch from Be p-doping for the base to C p-doping in the Rockwell process (which had already taken place). This permitted the increase of the dopant redistribution emitter current density limitation from 0.5 mA per square micron of emitter area to 1.0 mA. This doubling of the critical current density then enabled substituting the smaller emitter device directly into existing circuits which had fixed the peak current into these emitters at 2 mA. Because the resistance's in the device were much smaller than external bias resistors, direct substitution could be performed without altering any external resistance's. The smaller emitter width dimension of 1.2 µm of width was also tested by Rockwell as a part of its 100 GHz process development effort.

This halving of the emitter area alone without a change in the width to length aspect ratio of this opening would have resulted in approximately a doubling of the extrinsic portion of the base resistance which is sensitive to the length of the perimeter of the emitter facing active base region. This is estimated since the extrinsic base resistance was approximately 4 times the intrinsic value, and the extrinsic portion is inversely proportional to the perimeter length of the emitter. Consequently every effort was taken in the shrinking process to lengthen the emitter edge. Long "skinny" rectangular emitters are then preferred in this regard because they maximize the perimeter of the emitter for its given area. This 1.2 by 1.7 square micron emitter rented the "middle" of the evolution of the HBT . The 1.2 micron evolution presents the current limit to making the emitter "skinny" because this is the current minimum feature size of the process. For comparison, the IBM SiGe HBT has an emitter of 0.35 µm by 1 µm giving a 3:1 aspect ratio at only 10% of the baseline HBT area.

Round emitters, which were also candidates suggested by Rockwell, have the least perimeter for the given area enclosed, although all of that perimeter would be accessible as active base-emitter region. Round emitters also would inefficiently underutilize the area of the base pedestal around the four corner "fillets," being a proverbial round peg in a square hole. To utilize a round emitter fully all of its perimeter would have to face active base edge. This would necessitate placing a via directly on top of the emitter to enter that contact from M2, while presenting the base contact on M1. This would have permitted more layout flexibility for the M2 to access the emitter, which would have had some subtle layout improvements in cell density. Offsetting these advantages was the likelihood that the M2-emitter via presents a yield risk. The minimum feature size of that via, together with the known thickness variability of the ILD directly above the emitter suggested that a this via might not "land" properly on the emitter consistently for with the round case. Additionally as the transistor shrinks in future scalings this would limit the emitter area to a minimum M1-contact via which would have to be fairly big.

Instead it is argued that both base and emitter contacts for the rectangular emitter stripe could enter from M1 or from a short strip of ohmic metal out to an M1 overlayer. These were known to work well from the point of view of yield. An experimental lightly loaded ring oscillator was submitted as a partial reticle exploration on a Science Center fab at Newbury Park with this intermediate transistor, but the results are not yet available.

Unfortunately the first attempt at shrinking the emitter did not provide an opposing face off the emitter to an active base region on the short "ends" of the emitter stripe (the 1.2 micron ends). The reason for not doing this was to avoid changing too many features in one device evolutionary step. Only the emitter area shrinkage was undertaken in this experiment.

The normal reason for this would be a large design rule violation between two M1 lines for lines connecting to the base and emitter, as the would be too close together. However, upon examining a set of exploratory HBT layouts from Rockwell a transistor was observed that utilized only ohmic metal to make a short connection to the emitter and base. This avoided the M1-M1 design rule violation and made an opposing face possible. Furthermore, the ohmic metal spacing could be made so small as to permit a much smaller base emitter spacing. This spacing could be as small as 0.4 microns, although technically no actual feature size would be submicron. Only this spacing would be submicron. This would require extreme layer to layer lithographic registration accuracy, but not necessarily better resolution.

A specific reference ring oscillator has been used to estimate the relative importance of reduction of various parasitics during this device redesign effort. These are summarized in the following table (all resistance's are in Ohms, all capacitances are femto Farads, and all times are in picoseconds):

33

	O	M	C1	C2	C3	K1	K2	K3	A	T
W x L	1.4 x 3	1.4 x 3	1.4 x 3	1.2 x 2	1.4 x 1.2	1.2 x 1.7	1.2 x 1.7	1.2 x 1.7	1.2 x 1.7	1.2 x 1.7
Re	14	35	15	21	45	60	45	45	60	60
Rb	76	39	76	99	130	110	110	110	70	70
Cb	16	27	36	28	19	14	14	14	14	14
Rc	39	85	39	53	40	85	85	53	70	70
Bf	1000	1000	194	194	238	194	194	194	194	194
Tf	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	1.2
Tr	350	490	521	488	467	391	383	366	345	286

Table 1. Comparison of original estimate of ring oscillator time with measured time, and with various other estimates for evolved HBT models.

In the Table 1, **O** is the originally supplied set of SPICE model parameters for the "50 GHz Baseline " process, **M** is the model fitted by Rensselaer to ring oscillators fabricated by Rockwell, and checked against S-parameter sets measured by Rockwell and provided to Rensselaer, **C1** is a subsequent model supplied by Rockwell, with **C2** and **C3** being smaller emitter area models, **K1** is a model for the middle evolved HBT layout with **K2** and **K3** representing different assumptions on the impact on Re and Rc of the shrink. The prediction of the effect of shrinking the emitter to 1.2 x 1.7 square microns on Rc and Re is more difficult than for Cb and Rb. Next, **A** represents the best estimate of the most aggressively scaled device layout, shrinking base emitter separations to 0.4 microns, moving the collector contact closer to the emitter, and starting from the worst case estimates for the **K** series. Finally the last model. **T**, assumes a thinned base for the **A** model to decrease the base transit time. It can be seen that the only model to come close to the original ring oscillator time estimate of 350 picoseconds is the **A** model. This is the speed which the ring oscillator would need to exhibit in order for the architecture chips as designed to perform at the speed required for a 1000 MIPS operation. This suggests that some very aggressive layout alterations are required to achieve the speed assumed throughout the whole design project. At the time of writing this final report the ring oscillator corresponding to the **K** series is being fabricated by donation or reticle space by K.C. Wang at Rockwell, and the more aggressive **A** ring oscillator is being fabricated on an HSCD reticle. Funding for the HSCD subcontract to Rensselaer has been terminated due to funding cutbacks at the prime contract level. Hence this extra fabrication has been in the form of a donation to Rensselaer by Rockwell in an effort to resolve this device speed problem.

The ring oscillator is large enough to obtain some minimal feedback on the impact on yield from the use of these more aggressive transistors.

II.2.E.Conclusions

The RPI test chip fabricated in 1993 showed sufficient yield to verify the standard cells, register file, and ALU circuitry. The chip showed no self oscillations and low jitter validating our differential logic design and use differential signal routing and embedded testing approach with standardized multi-channel ceramic probes for testing at microwave frequencies. However, circuits with more than a few hundred devices had low yield. While some LFSR circuits worked at up to 2.3 GHz the test circuits were 33-50 % slower than expected. Based on device S-Parameter measurements and Rockwell's frequency dividers fabricated on the same reticle we concluded together with Rockwell that the device performance on this run was off, the maximum F_t of the HBTs was only 33 GHz rather than 50 GHz.

The HSCD reticle fabricated in 1994 contained three RPI chips and a passive test chip designed by RPI under an HSCD subcontract to Rockwell. The new stepper Rockwell had introduced clearly improved yields. Our VCO circuit performed at 13.66 GHz, but performance was still 33 % slower than expected based on SPICE simulations backannotated with a novel 3-D capacitance extractor. Other circuits and ringoscillators on the 'passive' test chip confirmed that the switching performance of the devices was slower than the predicted by Rockwell's SPICE model. However, S-Parameter measurements both at Rockwell and Mayo showed that the devices have indeed a maximum F_t of 50 GHz. Our investigation showed that the model incorrectly models switching device performance. The switching performance is dominated by the F_t of the device at low current levels, and not maximum F_t.

The measurements of capacitance test structures on the passive test chip revealed that the interlayer dielectrics are thinner than expected based upon the design manual. In large area parallel plate capacitors the M1-M2 dielectric is only 1.1 μm instead of 1.6 μm. Measurements of M1-M2 crossovers showed that the dielectric is only 0.9-0.95 μm thick indicating that the polyimide dielectric is not planarizing as well as it should. We have shrunk the width of local interconnects to compensate for the thinner dielectric layers taking advantage of a recent process upgrade.

Further, working in conjunction with Rockwell, we are currently exploring new switching devices that have smaller emitter sizes taking advantage of the doubling of the maximum emitter current after Rockwell switched from Be to carbon doping. The smaller emitter and base pedestal area lowers junction capacitances, increases the current density in the emitter so that maximum F_t is reached at lower current levels and thus improves switching performance. Several RPI test circuits with new devices are currently in fabrication. The new devices are drop in replacements for the devices used in our architecture reticle. Hence, the architecture reticle can be upgraded very quickly once we know which of the new devices meets or exceeds the switching performance of the model used for our designs and can be fabricated with sufficiently high yield.

35

III. LIST OF ALL PUBLICATIONS AND TECHNICAL REPORTS

[1]   ``Cell Library for Current Mode Logic using an Advanced Bipolar Process," (J. F. McDonald, H. J. Greub, T. Yamaguchi, and T. Creedon), I.E.E.E. J. Sol. State Cir., Special issue on VLSI, (D. Bouldin, guest editor), I.E.E.E. Trans. on Solid State Circuits, Vol. JSSC-26(#5), pp. 749-762, May, 1991.

[2]   ``F-RISC/I: Fast Reduced Instruction Set Computer with GaAs H-MESFET Implementation," Proc. I.E.E.E. Int. Conf. on Computer Des., (J. F. McDonald, C.K. Tien, C.C. Poon, H. Greub) Boston, MA, (I.E.E.E. Cat. # CH3040-3/91/0000/0293), pp. 293-296, October 14-16, 1991.

[3]   ``F-RISC/G: AlGaAs/GaAs HBT Standard Cell Library, "Proc. I.E.E.E. Int. Conf. on Computer Des., (J. F. McDonald, K. Nah, R. Philhower, J.S. Vanetten, S. Simmons, V. Tsinker, Maj. J. Loy, and H. Greub), Boston, MA, (I.E.E.E. Cat. # -3/91/0297), pp. 297-300, October, 1991.

[4]   ``Wideband Wafer-Scale Interconnection in a Wafer Scale Hybrid Package for a 1000 MIPS Highly Pipelined GaAs/AlGaAs HBT Reduced Instruction Set Computer," Proc. 1992 Int. Conf. on Wafer Scale Integration, ICWSI-4, San Francisco, January 20, 1992, Reprinted Hardbound by Computer Science Press, V. K Jain, and P. W. Wyatt, Eds. [I.E.E.E. CS#2482], pp. 145-154. (J. F. McDonald, R. Philhower, J. S. Van Etten, S. Dabral, K. Nah, and H. Greub).

[5]   ``Bypass Capacitance for WSI/WSHP Applications," Proc. Fifth Int. Conf. on WSI, ICWSI93, San Francisco, CA, M. Lea, Ed., I.E.E.E. Computer Soc. Press, pp. 218-228, February, 1993 (J. F. McDonald, H. Greub, R. Philhower, J. Van Etten, K. S. Nah, P. Campbell, C. Maier, Lt. C. J. Loy, P. Li, L. You, and T..-M. Lu).

[6]   ``Fluorinated Parylene as an Interlayer Dielectric for Thin Film MultiChip Modules,'' spring 1992 meeting of the Materials Research Society, Reprinted in Vol. 264 of the MRS Symposium Proceedings, Electronic and Packaging Materials Science VI, Paul S. Ho, K. A. Jackson, C.-Y. Li and G. F. Lipscomb, Eds., pp. 83-90, 1993 (J. F. McDonald, S. Dabral, X. Zhang, W. M. Wu, G.-R. Yang, C. Lang, H. Bakhru, R. Olsen and T.-M. Lu)

[7]   ``A 500ps 32 X 32 Register File Implemented in GaAs/AlGaAs HBT's,'' Proc. I.E.E.E. GaAs Symposium [I.E.E.E. Cat. 93CH3346-4], San Jose, Oct. 1993, pp. 71-75, (J. F. McDonald, K. S. Nah, R. Philhower, H. and Greub).

[8]   ``F-RISC/I: A 32 Bit RISC Processor Implemented in GaAs H-MESFET Super Buffer Logic,'' Proc. I.E.E.E. GaAs Symposium [I.E.E.E. Cat. #93CH3346-4], San Jose, CA, Oct. 1993, pp. 145-148, (J. F. McDonald, C.K. Tien, K. Lewis, R. Philhower, and H. J. Greub).

[9]   ``Frequency Domain (1kHz-40GHz) Characterization of Thin Films for Multichip Module Pakaging Technology,'' (J. F. McDonald, W.-T. Liu, S. Cochrane, X.-M. Wu, P.K. Singh, X. Zhang, D. B. Knorr, E. J Rymaszewski, J. M. Borrego, and T.-M. Lu), Elect. Lett., Jan. 20, 1994, Vol. 30(#2), pp. 117-118.

[10]   `Poly-tetrafluoro-p-xylylene as a Dielectric for Chip and MCM Applications,'' (J. F. McDonald, S. Dabral, G._Y. Yang, X. Zhang, & T.-M. Lu, J. Vac. Sci. and Technol., B 11(#5), Sept./Oct. 1993, pp. 1825-1830.

[11]   ``Application of a Floating-Random-Walk Algorithm for Extracting Capacitances in a Realistic HBT Fast-RISC RAM Cell.'' (J. F. McDonald, Y. L. Le Coz, R. B. Iverson, H. J. Greub, P. M. Campbell, and J. F. McDonald), Proc. I.E.E.E. VLSI Multi-Layer Interconnect Conf., V-MIC94, Santa Clara, CA, June, 1994, pp. 542-544.

[12]   ``Design of a Package for a High Speed Processor Made with Yield Limited Technology,'' (J. F. McDonald, A. Garg, J. Loy, and H. Greub), Proc. I.E.E.E. Fourth Great Lakes Symposium on VLSI, March 4-5, 1994, Notre Dame University, Indianna, [I.E.E.E. Cat. #94TH0603-1, Comp. Soc. # 5610-02], pp. 110-113.

[13]   ``Wiring Pitch Integrates MCM Wiring Domains,'' (J. F. McDonald, J. Loy, A. Garg, M. Krishnamoorthy), Proc. I.E.E.E. Fourth Great Lakes Symposium on VLSI, March 4-5, 1994, Notre Dame University, Indianna, [I.E.E.E. Cat. #94TH0603-1, Comp. Soc. # 5610-02], pp. 110-113.

[14]   ``Differential Routing of MCMs — CIF: The Ideal Bifurcation Medium,'' (J. F. McDonald, J. Loy, A. Garg, M. Krishnamoorthy), Proc. I.E.E.E. Int. Conf. on Computer Des., Cambridge, MA, [I.E.E.E. Cat. # 94CH35712], pp. 599-603, October 10-12, 1994.

[15]   ``Thermal Design of an Advanced Multichip Module for a RISC Processor,'' (J. F. McDonald, A. Garg, J. Loy, H. Greub, T. -L. Sham), Proc. I.E.E.E. Int. Conf. on Computer Des., Cambridge, MA, [I.E.E.E. Cat. # 94CH35712], pp. 608-611, October 10-12, 1994.

[16]   ``Three Dimensional Stacking with Diamond Sheet Heat Extraction for Subnanosecond Machine Design,'' (J.F. McDonald, H. Greub, A. Garg, P. Campbell, S. Carlough, C. Maier), Proc. 1995 Int. Conf. on Wafer Scale Integration, ICWSI-7, San Francisco, January 20-22, 1995, Reprinted in Hardbound by Society Press, S.K. Tewksbury and S. K. Tewksbury, and G. Chapman, Eds. [I.E.E.E. CS #2482], pp. 62-71.

[17]   ``Design of a 32-bit Monolithic Microprocessor Based on GaAs H-MESFET Technology,'' in review for I.E.E.E. Transactions on VLSI Systems,'' (J. F. McDonald, C.-K. V. Tien, K. Lewis, H. J. Greub, and T. Tsen).

IV. LIST OF ALL SCIENTIFIC PERSONNEL SHOWING ADVANCED DEGREES EARNED BY THEM WHILE EMPLOYED ON THE PROJECT

[1]   Lt. Cmdr. James Loy, "**Differential Routing Tools for High Speed GaAs HBT CML Circuits,**" Ph.D. 1993.

[2]   Robert Philhower, "**Spartan RISC Architecture for Yield Limited Technology,**" Ph.D. 1993.

[3]   Kyung Suc Nah, "**An Adaptive Clock Deskew Scheme and a 500 ps 32 by 8 Bit Register File for a High Speed Digital System,**" Ph.D. 1994.

[4]   C.-K. Vincent Tien, "**System Design, Analysis, Implementation and Performance Evaluation of a 32 Bit RISC Processor Based on GaAs HMESFET Technology,**" PhD 1994.

V. LIST OF INVENTIONS BY NAME

No formal patent applications have been filed during this grant due to lack of funds for legal expenses. However, it is possible that the ideas presented in Appendix C on clock deskew circuitry could qualify for a patent if one were to be submitted.

VI. BIBLIOGRAPHY

[1]   C.Y. Chang, and Francis Kai, GaAs High-Speed Devices, John Wiley, 1994.

[2]   R. Anholt, Electrical and Thermal Characterization of MESFETs, HEMTs, and HBTs, Artech House, 1995.

[3]   D. J. Roulston, Bipolar Semiconductor Devices, McGraw Hill, 1990.

[4]   B. Jalali, and S. J., Pearton, Eds., InP HBTs, Growth, Processing and Applications, Artech House, 1995

[5]   R. Williams, Modern GaAs Processing Techniques, Artech House, 1991.

[6]   U. Ciligiroglu, Systematic Analysis of Bipolar and MOS Transistors, Artech House, 1994.

[7]   F. Ali, and A. Gupta, Eds., HEMTs & HBTs, Artech House, 1991.

[8]   J. W. Mayer and S.S. Lau, Electronic Materials Science for Integrated Circuits in Si and GaAs, Macmillian, 1990.

[9]   N. Kanopoulos, Gallium Arsenide Digital Integrated Circuits, Prentice Hall, 1989.

[10]   S. Long, and S. Butner, Gallium Arsenide Digital Integrated Circuit Design, McGraw Hill, 1990.

[11]   V. Milutinovic, Ed., Microprocessor Design for GaAs Technology, Prentice Hall Advanced Reference Series in Engineering, 1990.

[12]   M. Katevenis, <u>Reduced Instruction Set Computer Architectures</u>, MIT Press, 1984.

[13]   J. R. Ellis, <u>Bulldog: A compiler for VLIW Architectures</u>, MIT Press, 1985.

[14]   S. S. Sapatnekar, and S.-M. Kang, <u>Design Automation for Timing Driven Layout Synthesis</u>, Kluwer Academic Publishers, 1993.

[15]   R. Jain, <u>The Art of Computer Systems Performance Analysis</u>, J. Wiley & Sons, 1991.

[16]   S. A. Przybylski, <u>Cache and Memory Hierarchy Design</u>, Morgan Kaufman Publishers, 1990.

[17]   H. B. Bakoglu, Circuits, <u>Interconnections, and Packaging for VLSI</u>, Addison Wesley Publishers, Inc., 1990.

[18]   D. A. Patterson, & J. L. Hennessy, <u>Computer Organization & Design - The Hardware Software Interface</u>, Morgan Kaufman Publishers, 1994.

[19]   E. J. Rymaszewski, <u>Handbook of Microelectronics Packaging</u>, Van Nostrand, 1990.

[20]   F. E. Gardiol, <u>Lossy Transmission Lines</u>, Artech House, 1987.

VII. APPENDICES

VII.1.   Appendix A

VII.1.A. High Speed Circuit Design (HSCD) Measurements

VII.1.A.1. HBT Test Wafer

A reticle containing test chips was submitted to Rockwell for fabrication in July 94. The layout of the reticle is shown in Figure 25. This reticle contains four RPI chips: passive test chip, standard cell test chip, 20 GHz voltage controlled oscillator (VCO) test chip, and register file test chip. The first fabricated wafers were received in December 94.
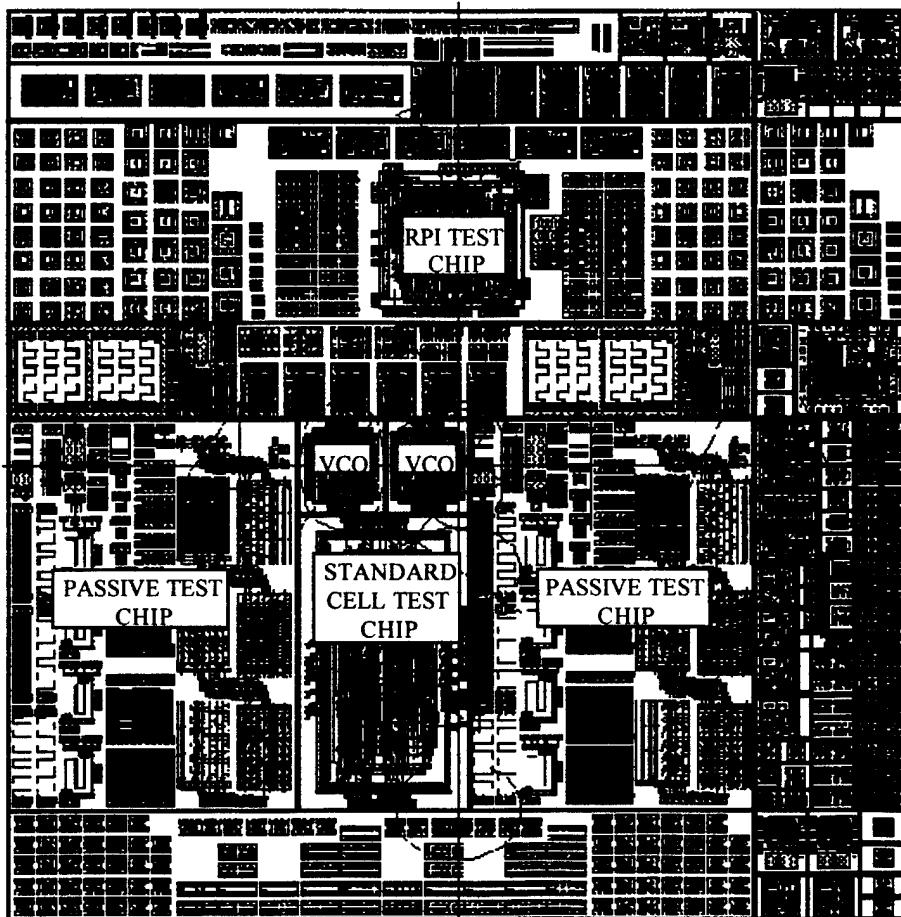
Figure 25. Layout of the RPI-Rockwell Reticle.

The mask contains a variety of circuits to determine the basic cell performance as a function of power supply voltage, current level, temperature and processing variations

43

Specifically, the passive test chip contains test structures to measure wiring parasitics on a HBT chip. It also carries ring oscillators and gate delay chains to provide basic delay information as a function of capacitive load and fanout. Other chips contain a number of key circuits used in the main architecture chips. The 20 GHz VCO chip has a high-speed voltage controlled oscillator on the chip with several other circuits to test the performance of the process. The register file test chip is an optimized version of the previous test chip fabricated at Rockwell. It also includes a high-speed carry chain macro and associated support circuits. The standard cell test chip contains a number of representative standard cells used in the F-RISC/G chips and tests the implementation of the boundary scan test scheme applied to test the instruction decoder and the datapath chips.

Ripple divider circuits are used to determine flip-flop performance. Several functional circuits are also used including a 2:1 mux, 1:2 demux, 4x4 parallel multiplier and a 7-bit LFSR. These circuits are used to evaluate yield and cell performance in a variety of conditions. Additional test structures were included to measure individual cell and device characteristics.

Currently, the passive test chip is being tested at RPI. The chip and the test results are described in the next few sections.

VII.1.A.2. Passive Test Chip

The layout of the chip is shown in Figure 26. This chip contains both the passive test structures and the active test structures.
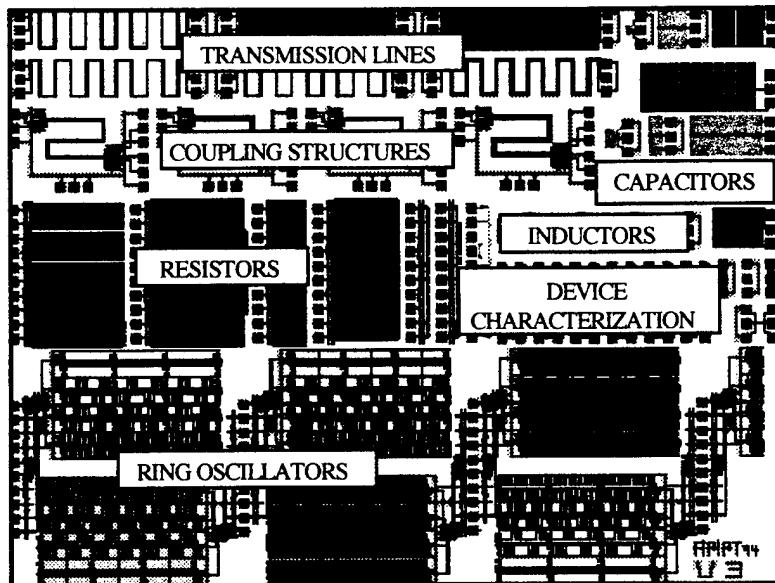
Figure 26. Layout of the Passive Test Chip.

The passive structures are meant for measuring wiring parasitics on a AlGaAs/GaAs HBT chip and comparing the measured results with results obtained from CAD tools. The structures are divided into five categories — capacitors, inductors, probe calibration, transmission lines, and resistors.

44

The active structures are divided into three categories — coupling, device characterization, and ring oscillators. The coupling structures allow measuring the coupling between differentially coupled wires and single-ended wires. A number of device-characterization structures are provided close to the ring oscillators to correlate the measurements with the device performance. The ring-oscillators are loaded with different interconnect capacitances to show the effect of capacitive loading on the wires. These oscillators are made up of standard Q1 and the new round Q1 transistors. The oscillation frequencies of these structures lie in the range of 0.5 GHz - 3.0 GHz.

VII.1.A.3. MIM Capacitors Test Results

MIM capacitors are made between M1 and M2 layers sandwiching only the nitride layer. There were two instances of these capacitors on the chip with a theoretical (based on the design rule manual) capacitance of 2.08 pF and 8.32 pF respectively. A series RLC model was fitted to the fabricated capacitors. The extracted capacitance showed as much as 10% lower capacitance than the predicted values as shown in Table 2 and Table 3.

Table 2. Structure 1 (Theoretical Capacitance = 2.08 pF)

Die	Site	Extracted R [ohm]	Extracted L [pH]	Extracted C [pF]	Difference (Theo. vs Ext.)
00	1	1.0187	80.8	1.923	-7.5 %
00	2	0.9747	83.1	1.936	-6.9 %
11	1	1.0352	82.8	1.960	-5.7 %
11	2	1.0048	84.5	1.983	-4.7 %
-11	1	1.0205	79.9	1.940	-6.7 %
-11	2	0.9795	79.0	1.952	-6.1 %
-1-1	1	0.9983	78.0	1.926	-7.4 %
-1-1	2	0.9563	80.9	1.918	-7.8 %
1-1	1	0.9860	77.2	1.944	-6.5 %
1-1	2	0.9920	80.0	1.940	-6.7 %
0-2	1	2.0857	73.4	1.118	-46.2 %*
0-2	2	0.9701	79.8	1.935	-6.9 %

*Wafer edge

Table 3. Structure 2 (Theoretical Capacitance = 8.32 pF)

Die	Site	Extracted R [ohm]	Extracted L [pH]	Extracted C [pF]	Difference (Theo. vs Ext.)
00	1	0.9051	77.0	7.48	-10.0 %
00	2	0.8815	79.4	7.52	-9.1 %
11	1	0.9224	78.7	7.63	-8.2 %
11	2	0.9232	81.5	7.70	-7.4 %
-11	1	0.9175	76.5	7.53	-9.4 %
-11	2	0.9483	76.1	7.58	-8.9 %
-1-1	1	0.8917	73.8	7.45	-10.4 %
-1-1	2	0.9026	77.7	7.45	-10.4 %
1-1	1	0.9114	73.8	7.53	-9.4 %
1-1	2	0.9230	77.1	7.50	-9.8 %
0-2	1	0.9236	72.5	7.75	-6.8 %
0-2	2	0.9302	76.6	7.50	-9.8 %

VII.1.A.4.Parallel Plate Capacitors Test Results

Parallel plate or overlap capacitors are made by overlapping interconnect metal layers. There were three M1/M2 parallel plate capacitors on the chip with a theoretical capacitance of 1.09 pF, 2.18 pF, and 5.18 pF respectively. The extracted capacitance showed as much as 45% higher capacitance than the predicted values as shown in the tables below.

Table 4. Structure 1 - M1/M2 (Theoretical Capacitance = 1.09 pF)

Die	Site	Extracted R [ohm]	Extracted L [pH]	Extracted C [pF]	Difference (Theo. vs Ext.)
00	1	0.4137	75.5	1.58	44.9 %
00	2	0.3875	77.1	1.56	43.1 %
11	1	0.3774	75.8	1.57	44.0 %
11	2	0.3983	78.5	1.57	44.0 %
-11	1	0.3634	74.3	1.58	44.9 %
-11	2	0.3866	74.5	1.58	44.9 %
-1-1	1	0.4015	72.6	1.58	44.9 %
-1-1	2	0.3451	75.8	1.58	44.9 %
1-1	1	0.3911	71.7	1.55	42.2 %
1-1	2	0.3991	74.0	1.55	42.2 %

*Wafer edge

Table 5. Structure 2 - M1/M2 (Theoretical Capacitance = 2.18 pF)

Die	Site	Extracted R [ohm]	Extracted L [pH]	Extracted C [pF]	Difference (Theo. vs Ext.)
00	1	0.4765	86.7	3.08	41.2 %
00	2	0.4352	88.6	3.07	40.8 %
11	1	0.4768	87.9	3.06	40.3 %
11	2	0.4367	90.1	3.06	40.3 %
-11	1	0.4741	85.3	3.08	41.2 %
-11	2	0.4530	85.4	3.08	41.2 %
-1-1	1	0.4785	84.1	3.12	43.1 %
-1-1	2	0.4168	87.1	3.12	43.1 %
1-1	1	0.4650	82.8	3.05	39.9 %
1-1	2	0.4175	85.3	3.05	39.9 %
0-2	1	0.4786	81.7	1.93	-11.4 %*
0-2	2	0.4157	84.9	3.09	41.7 %

*Wafer edge

Table 6. Structure 3 - M1/M2 (Theoretical Capacitance = 5.18 pF)

Die	Site	Extracted R [ohm]	Extracted L [pH]	Extracted C [pF]	Difference (Theo. vs Ext.)
00	1	0.5570	85.2	7.12	37.4 %
00	2	0.5571	87.6	7.07	36.4 %
11	1	0.5771	85.9	7.13	37.6 %
11	2	0.5608	88.5	7.08	36.6 %
-11	1	0.5706	85.0	7.11	37.2 %
-11	2	0.5436	84.8	7.09	36.8 %
-1-1	1	0.5805	82.2	7.14	37.8 %
-1-1	2	0.5519	85.9	7.13	37.6 %
1-1	1	0.5852	82.0	7.02	35.5 %
1-1	2	0.5725	84.0	7.01	35.3 %
0-2	1	0.5591	80.8	7.73	49.2 %*
0-2	2	0.5560	84.8	7.07	36.4 %

Table 7. Difference between measured and expected values of plate capacitors

Parallel Plate Cap Type	Size [μm]	Meas Cap [fF]	Expected Cap [fF]	Difference
M2/M3	250 x 160	725	467	+55.0%
M1/M2	250 x 160	858	606	+41.5%
M1/M3	250 x 640	1462	1055	+38.5%

VII.1.A.5.Resistors

These structures are designed to investigate the effect of the line width, corners, and processing steps on resistance's. The results are summarized in Table 8 . All the sheet resistanceŌs (M1,M2,M3,NICR,WSIN) were found to agree with the Rockwell specifications (or better) except the WSIN resistors which were within 15%.

Table 8. Interconnect sheet resistance measurements

| No	Resistor Type	Measured Sheet Resistance [ohms/sq]		Theoretical Sheet Resistance [ohms/sq]	
		Mean	Std. Dev.	Mean	Std. Dev.
1	M1	0.055	0.00039	0.055	0.0036
2	M1 (thru collector contacts)	0.062	0.00057		
3	M2	0.0173	0.00019	0.025	0.0020
4	M2 (orthogonally loaded with M1)	0.0176	0.00033		
5	M2 (Maximally loaded with VIA 12)	0.0190	0.00038		
6	VIA12	0.0395	0.00028		
7	M3	0.0144	0.00012	0.015	0.0004
8	M3 (orthogonally loaded with M1)	0.0145	0.00012		
9	M3 (orthogonally loaded with M2)	0.0159	0.00022		
10	M3 (on top of devices)	0.0144	0.00011		
11	M3 (maximally loaded with VIA 23)	0.0152	0.00017		
12	VIA 23	0.0199	0.00038		
13	NiCr	48.985	1.768	51.4	1.4
14	WSiN	253.5	14.13	290.5	8.23

From 1 and 2 it can be seen that any connection through a collector increases resistance. From 3 and 4, M2 has a higher sheet resistance when drawn orthogonally on top of M1 wires. From 3, 4, and 5, M2' sheet resistance increases with VIA12 in the path. M3 sheet resistance goes up if it is drawn orthogonally on top of M2 wires (from 7 and 9).

VII.1.A.6. Ring Oscillator Test Results

As HBT design is almost always designed with differential logic it was felt that loaded ringoscillator with several of these differential line configurations should also be included on the 'passive' test chip. These structures include wires with varying nearby grounded conductors, wires with adjacent differential lines, wires with metal planes on other layers, signal line overcrossings etc. To address difficulties in measuring the parasitics directly these structures were incorporated into ringoscillator circuits which could be simulated with SPICE using the extracted capacitances provided by tools such as METAL by OEA and QuickCAP by RLC, and then comparing the frequency of oscillation between the calculated waveforms and measured waveforms.

Since structures described above involve some active transistor devices, a means for measuring these device characteristics in the same general vicinity on the wafer and die are provided with special probe de-embedding sites to characterize the HBT's located in that area. There are deembedded transistors and deembedded Schottky diodes on the chip.

Figure 27 shows a plot between the measured sixteen-stage ring oscillator delay and the load capacitance at the output of each stage. The measured delay was found to be more than the simulated delay based on the capacitance extracted from layout and 50 GHz process design rules. The *Rockwell-50* and *Rockwell-w2* curves show the expected behavior of the oscillator. The *Rockwell-33* curve shows the behavior of a 33 GHz process based on the results obtained from an earlier wafer run. The *C=1.4* curve shows the oscillator behavior assuming a 50 GHz process with a 40% increase in the load capacitance due to reduced dielectric thickness. The measured results are approximated very well assuming a 33 GHz process and a 40% increase in the interconnect capacitance as shown by the *Rockwell-33, C=1.4* curve.
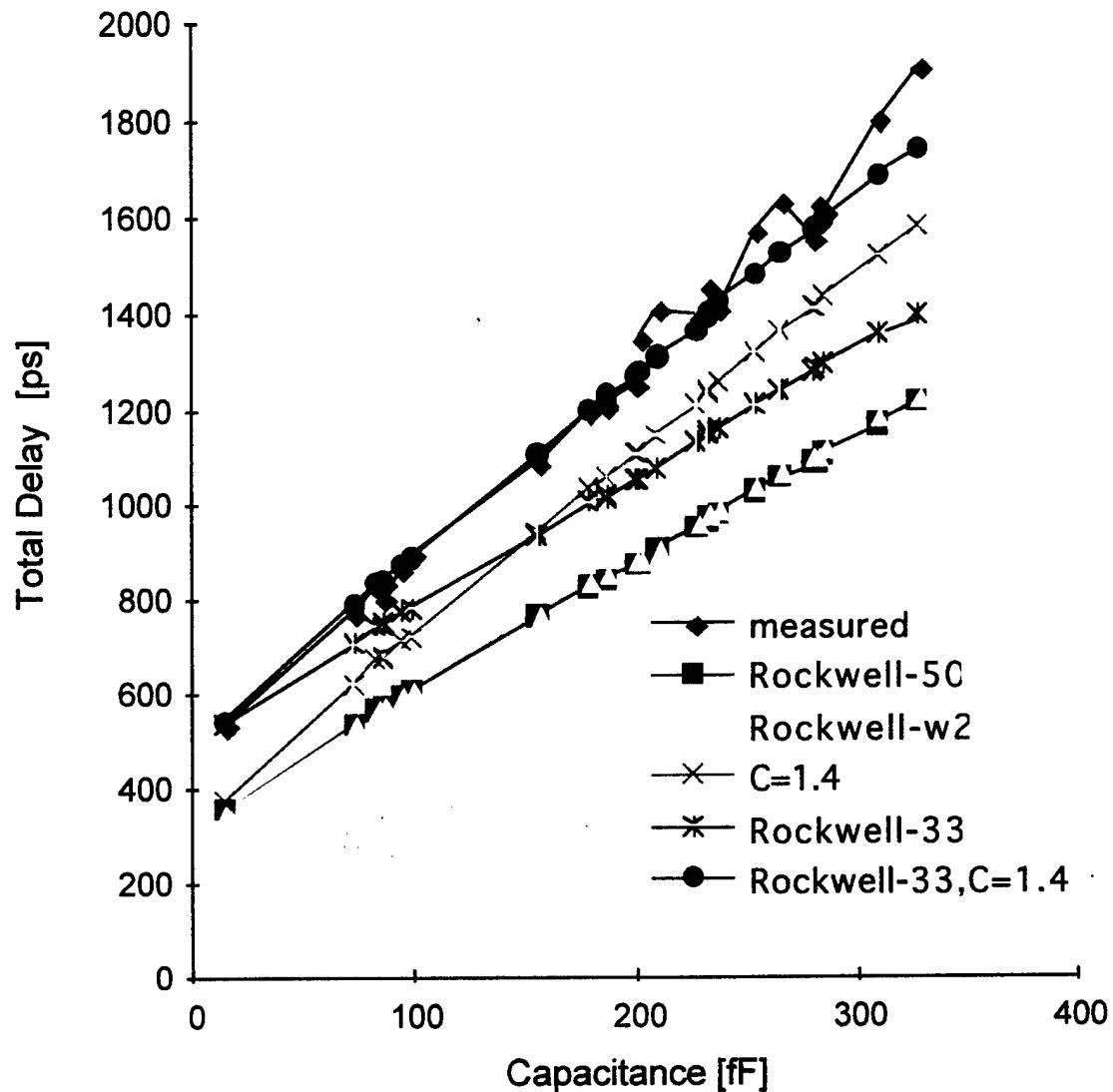
Figure 27. Ringoscillator Delays on RPI passive Test Chip (Total Delay = delay through sixteen stages, Capacitance = estimated load capacitance at each stage).

VII. 2. A. Optimization of the Register File used in the RPI Testchip and Datapath Chip

After the modifications to the memory cells and the address decoders were completed (as described in the last semiannual report), simulations with PSPICE (which included the wiring capacitances extracted with our new 3-D capacitance extractor) revealed that the register file was still too slow. In order to improve the access time, other cells were examined using the QuickCap capacitance extraction tool. As a result, the threshold voltage generator, address-line drivers, read-write logic and sense amplifiers were modified. In addition, the availability of a third level of metal opened up new layout possibilities which were explored and integrated into the optimized register file.

Figure 28 depicts the location of the changes within the register file. These changes are described below.

Most of the changes were made possible by the recent process upgrade to a third level of metal which could be routed over devices. This allowed the designer to produce layouts with less capacitance and more symmetry, thereby improving the circuit speed while reducing skew within a differential signal pair. Because the register file is an analog circuit which is highly sensitive to capacitance, symmetry in layout is critical. Based upon experience with the 20 GHz "Challenge" Chip, the designer of the VCO was selected to redesign the register file. Because the register file was already incorporated into two other layouts, it was also extremely important to maintain the original signal input/output locations. Although this constraint was always met, it did reduce the symmetry of the layout.

VII.2.A.1. Threshold Voltage Generator

There were a number of reasons for optimizing this circuit. Most of all, parts of this circuit must match exactly with the layout and orientation of both the memory cell and the wordline pullup resistors, hence the optimization of the memory cells dictated the redesign of the Threshold Voltage Generator. Other justification came from the use of a two-level metal process for the original design. As a result, the layout was unnecessarily complex for use with a three-level metal process, therefore it was decided that the circuit would be redesigned from scratch in order to fully utilize the new process. This new layout also allowed the use of monolithic microwave integrated circuit (MMIC) capacitors, and as a result, the overall size of the layout was reduced considerably.

Read/Write Line Buffer

Read/Write Logic Blocks

Threshold Voltage Generator

Address Line Drivers

Power Rail Metallization

Address Lines

Address Line Decoders

Memory Cells

Sense Amplifiers
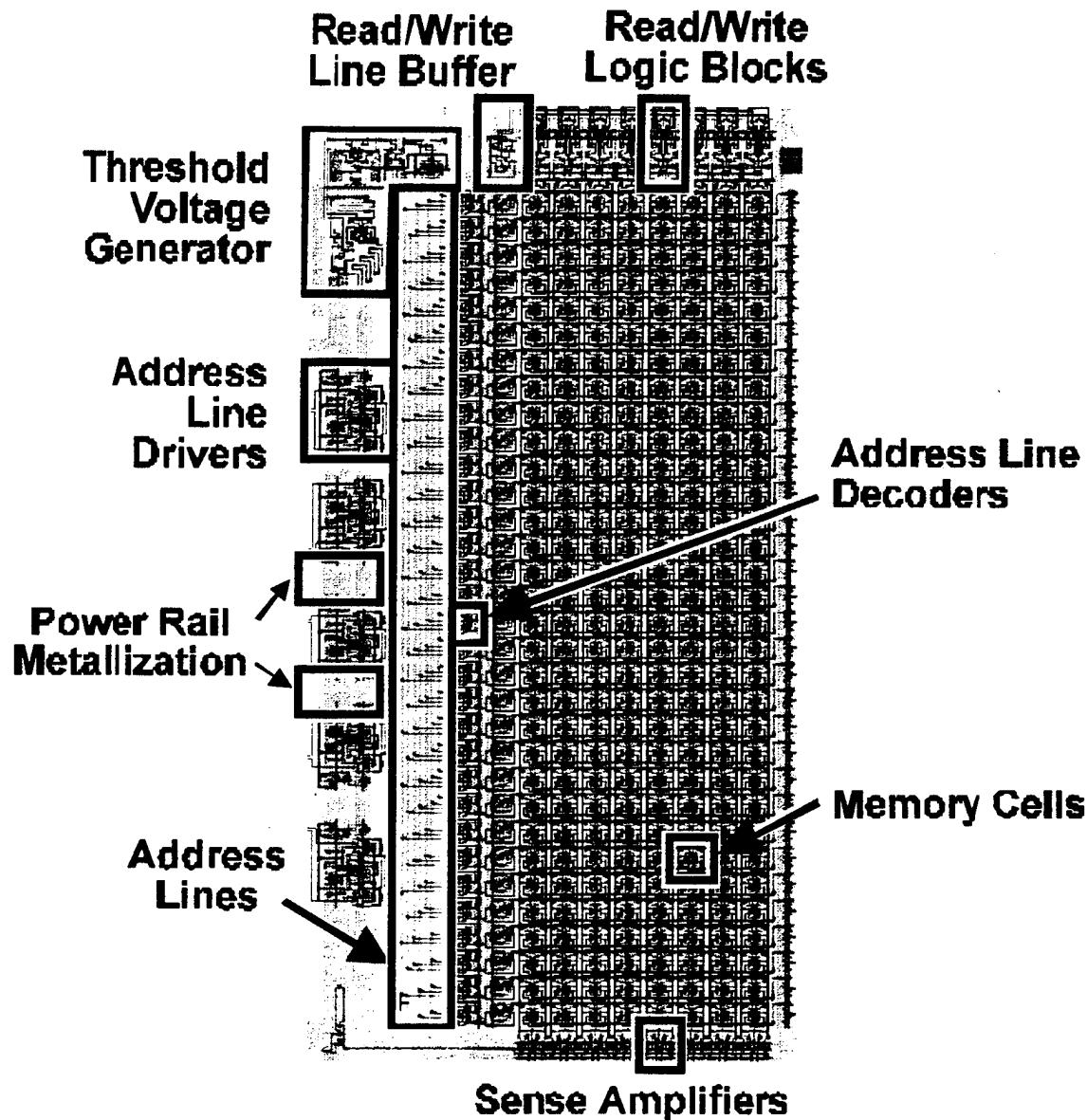
Figure 28: Register file modifications

VII.2.A.2. Address Line Drivers

As with the Threshold Voltage Generator, the original Address Line Driver was designed for a two-level metal process, resulting in a dense, asymmetrical layout with high parasitic capacitance. In order to efficiently utilize the new process, this circuit was also redesigned from scratch. Drawing upon experience with the high-speed VCO, the design methodology focused explicitly upon creating balanced, symmetric signal paths to ensure matched delay. As a result, the new optimized layout was significantly smaller than the original design. The savings in area were transferred to reducing capacitance on adjacent address lines by increasing the spacing between lines and between the driver and the lines. The Address Line Driver optimization was constrained by the original position of the register file input connections.

VII.2.A.3. Power Rail Metallization Changes

In optimizing the Address Line Drivers, it became possible to optimize the power rails within the register file. The original design required several alternating power and ground connections to the address driver side of the chip simply because a power connection placed between two address line drivers could not be extended beyond those two cells. By placing the power and ground rails in the third level of metal, the rails may be routed over the cells and thus all drivers may share the same supply rails. This helps reduce voltage droop along the rails and allows more flexibility in providing power to the register file macro.

VII.2.A.4. Address Line Metallization Changes

The Address Line Drivers are used as a buffer between the register file address line inputs and the internal address lines. The internal lines run the height of the macro and are connected to the 32 address line decoders. Crossover capacitance on the internal address lines can be significant and should be minimized, hence the metallization scheme was modified to take advantage of the third level of metal . By changing the address lines from metal2 to metal3, the crossover capacitance between the decoder inputs and the address lines was significantly reduced.

VII.2.A.5. Sense Amplifier Changes

The Sense Amplifiers were modified in order to reduce crossover capacitance and increase drive current capabilities. The internal supply rails were rerouted over devices using metal3 and the VSS rail was split into two rails in order to reduce capacitance. The drive current was boosted by replacing a normal Q1 transistor with a high-current Q3 device. The Sense Amplifier optimization was constrained by the original position of the register file output connections.

VII.2.A.6. Addition of Read/Write Buffer

A buffer was added to the Read/Write input signal to drive the eight Read/Write Logic cells. This buffer reduced the loading on the input signal and thus improved the access time of the register file. The addition of the buffer was made possible by the reduced area of the redesigned threshold voltage generator cell. The Read/Write Buffer placement and routing was constrained by the original position of the register file input connections.

VII.2.A.7. Read/Write Logic Changes

The Read/Write Logic was also optimized to take advantage of the third level of metal. Power rails were repositioned within the cell in order to reduce capacitance. In addition, the circuit was redesigned to remove a device and improve symmetry between the signal paths. The Read/Write Logic optimization was constrained by the original position of the register file input connections.

VII.3.A. Clock Distribution

The clock distribution of subnanosecond clock signals on an MCM is difficult since even relatively small amounts of skew can make up a significant fraction of the short clock cycle. For example, if data is transferred synchronously between two chips on the MCM within a 500 ps cycle and the clock skew is 50 ps only 400 ps are available for the transfer in the worst case. In addition, there will be skew in the on-chip clock distribution tree that provides the clock for the input and output latches on the two chips which can further reduce the available data transfer time. Thus a low skew clock distribution scheme on the MCM and on the chips is essential for subnanosecond computers.

We have developed a clock distribution scheme with active skew compensation based on digital delay lines and Phase Locked Loops (PLL). The skew compensation scheme can compensate for slowly varying delays due to temperature effects or water take-up, a problem with Polyimides. A test chip has been designed, laid out, and verified for evaluation of the clock distribution scheme at 2 GHz. The test chip contains several additional features to measure clock jitter and to increase testability and observability of key control signals.

Figure 29 shows the clock distribution scheme. A clock distribution chip provides a clock distribution channel for each clocked chip on the MCM. Each channel is essentially a PLL clock loop. The master clock is sent through a digital delay line on the forward path through a clock driver over a MCM transmission line to a clocked chip. The clocked chip receives the clock signal and feeds it to its four phase clock generator and returns the clock signal back to the clock distribution chip on a matched transmission line. The clock distribution chip receives the clock return signal and sends it through a matched digital delay line to the phase detector of a PLL controller. The controller will adjust the control voltage of the digital delay lines such that the phase difference or phase error between the master clock and the clock return signal is zero. In the ideal case all delays on the forward and return path are exactly matched and the clock arrives at the four phase generator on the receiving chip at $0.5* n * T_{clk}$ if the clock loop round trip delay is $n*T_{clk}$ and the PLL is in lock. Once all N clock channels are in lock, each receiving chip receives the master clock with a delay of $0.5*n*T_{clk}$ if we constrain the delays on each clock channel such that the clock delay multiplier n is the same for all clock channels.

The clock distribution chip contains further a system startup controller that generates the Sync signal that synchronizes the four phase generators on the receiving chips. The four phase generator switches to the next phase at every clock signal transition, thus a clock phase is only 250 ps long. Without synchronization the clocked chips might receive the clock without skew, but be in a different phase. The master clock must be stopped for a clock period in order to distribute the Sync signal to all receiving chips since the 250 ps delay between clock transitions is not sufficient to distribute the Sync signal to all chips on the MCM.

In order to prevent the clock loops from locking with different clock delay multipliers the following conditions must be met:

$$max(Delay\_of\_Delay\_Line) + max(Transmission\_Line\_Delay\_Missmatch) < T_{clk}$$

$$min(Delay\_of\_Delay\_Line) - max(Transmission\_Line\_Delay\_Missmatch) > -T_{clk}$$
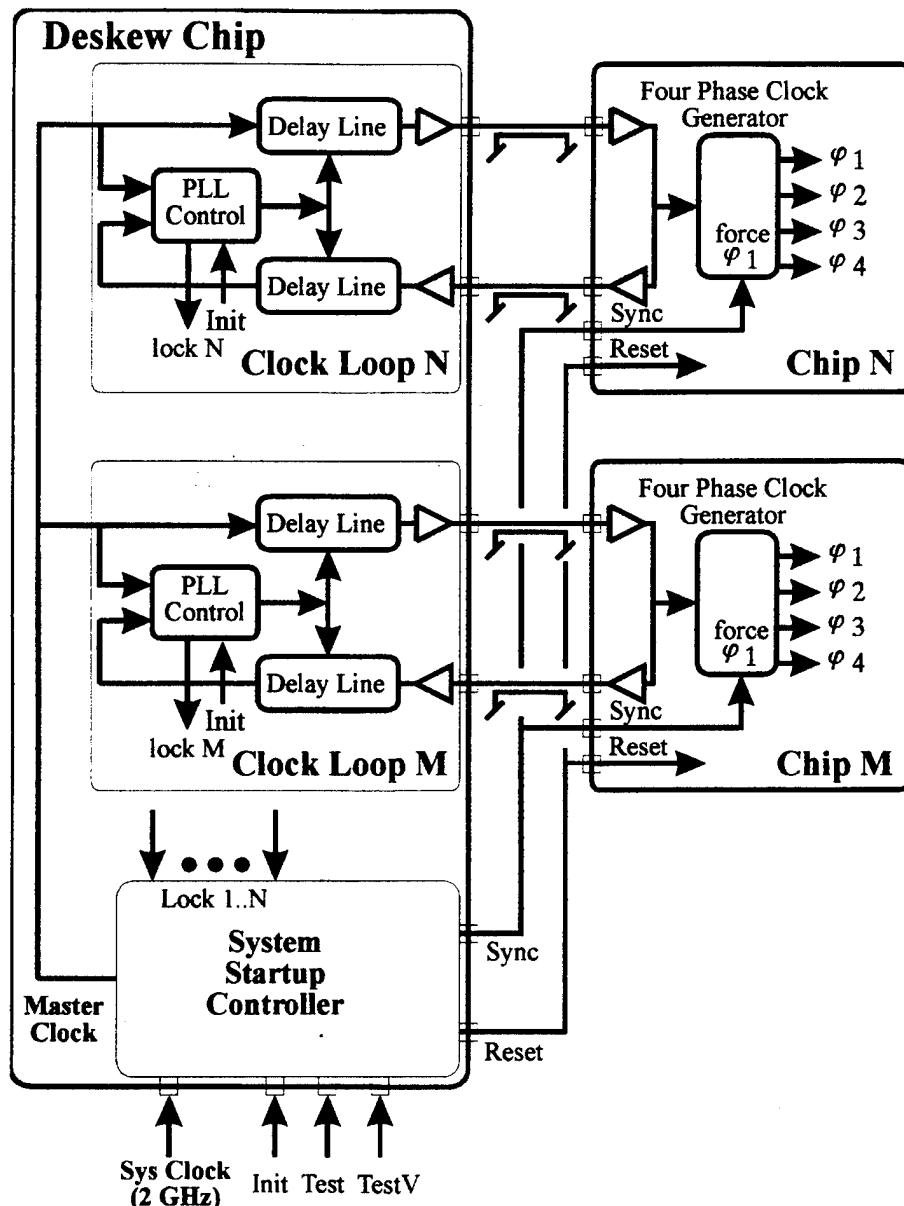
Figure 29. Active Clock Skew Compensation.

The maximum delay of the digital delay lines with respect to the initial delay, the Init signal forces the delay control signal to zero, is 125 ps and the minimum delay is -125 ps, thus the maximum tolerable delay mismatch between the clock distribution channels must be below 125 ps for a 2 GHz clock signal.

VII.3.B. Phase Locked Loop Controller

The phase locked loop controller adjusts the control voltage of the digital delay lines such that the phase difference between the master clock and the return clock is zero and the PLL stays in lock even if the interconnect or driver/receiver delays vary slowly. The controller is more complicated than in a PLL for frequency control since no VCO is present and some of the non-ideal behavior of phase detectors becomes important. The phase difference or phase error is measured with the three state phase detector shown in Figure 30. The phase detector has actually a fourth state (11) with both output signals UP and DOWN high simultaneously. If the phase detector is in state (11) it gets cleared by the AND gate after the propagation delay through the AND and the Reset delay of the master slave latch. If one of the input signals (V, R) goes through a positive transition while the phase detector is in state (11) or the clear signal is still active the transition gets lost and the phase detector switches characteristics. The two characteristics of an ideal three state phase detector are shown in. The switch will happen as soon as the phase difference is outside of the permissible phase range of the phase detector. The characteristics are offset by one clock cycle.
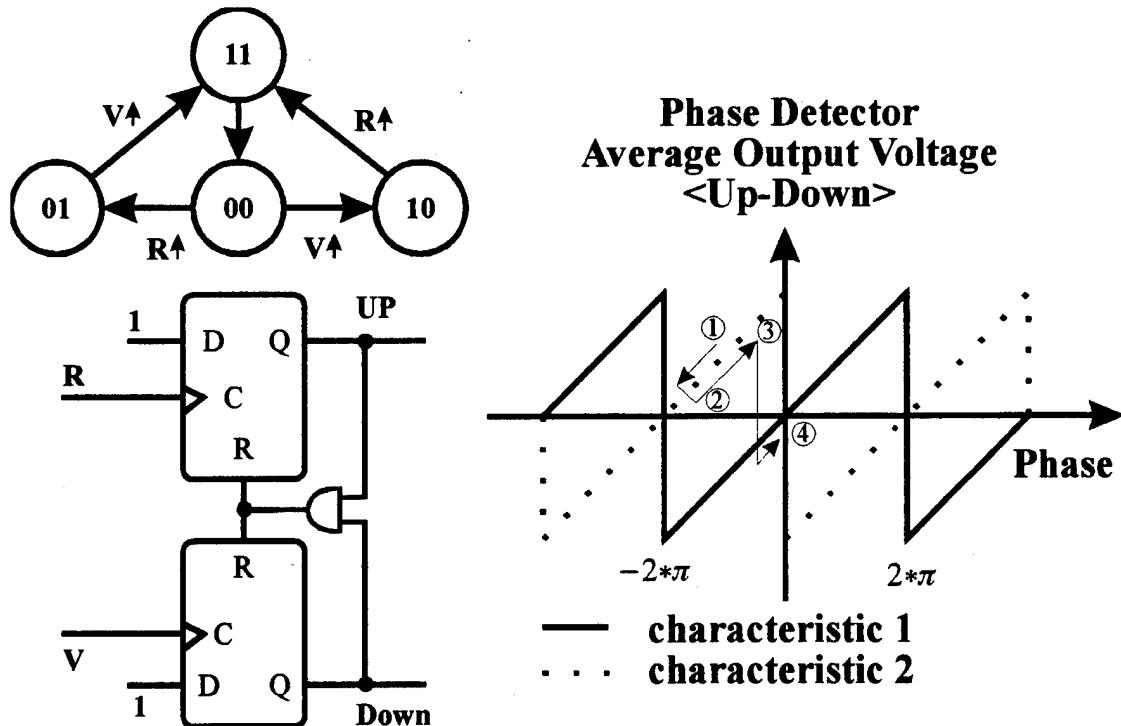

Figure 30. Three State Phase Detector.

Figure 31 shows the HBT phase detector characteristic for a 2 GHz clock signal. The trace shows the averaged phase error signal. The actual phase error signal generated from the Up, Down signals of the phase detector is a positive or negative pulse train. The actual

phase range is only $-\pi$ to π instead of the $-2*\pi$ to $2*\pi$ range of the ideal phase detector even though the latches have been optimized for a fast reset.

It is important to note that the sign of the phase error signal changes if the phase detector switches characteristics. Which characteristic the phase detector is on when the PLL starts up depends on initial conditions. Since the phase detector can be on characteristic 1 or 2 when the PLL starts up the error signal generated from the UP, DOWN signal for the PLL can have either sign!
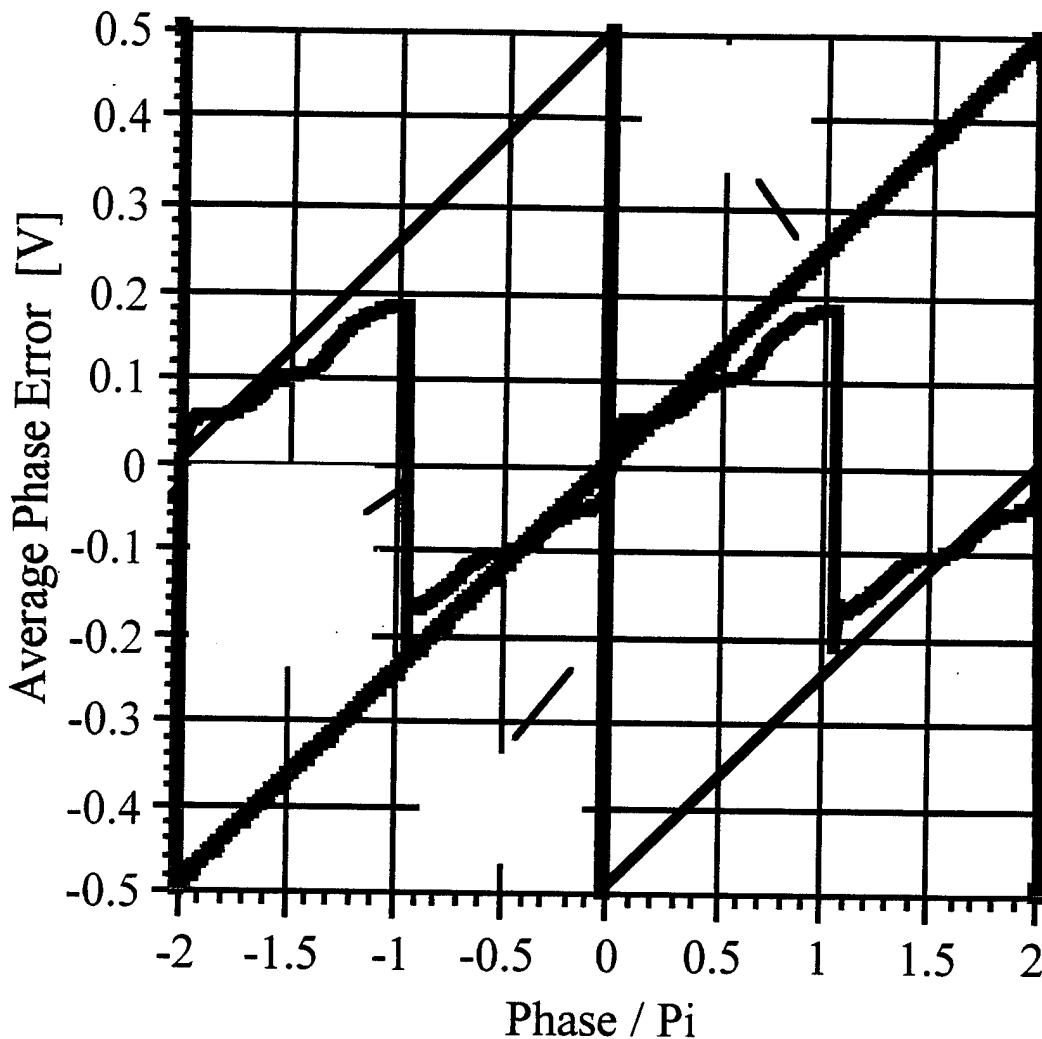
Figure 31. HBT Phase Detector Characteristic.

If the phase detector comes up in the wrong state or characteristic the PLL will have positive feed back and drive the PLL output voltage to its upper or lower limit, the PLL latches up! The controller must detect this situation and force the phase detector to change to the other characteristic. Unfortunately the phase detector is close to a zero of the current characteristic and the phase difference will be out of the range for the characteristic that we would like to switch to. Thus the phase detector will switch right back to the characteristic that lead to the latch up. An indirect approach must be taken to force a switch to the characteristic that provides negative feedback.

Figure 33 shows the PLL controller needed for each clock distribution channel. If the phase detector is on the wrong characteristic when the PLL starts up (situation 1 in Figure 30) the controller detects a PLL latch up with the two comparators that check whether the loop filter output voltage has reached the upper or lower voltage limit (situation 2). The loop filter has been replaced with an integrator to increase loop gain and reduce the steady state error of the PLL. If either limit is reached the corresponding comparator sets a latch that will force the Up, Down signal converter to output either high or low voltage. This will drive the phase difference outside of the range of the current phase detector characteristic and thus force a change over to the characteristic that provides negative feedback. The change in sign is detected by a novel differential Schmitt Trigger circuit which will reset the latch (situation 3).
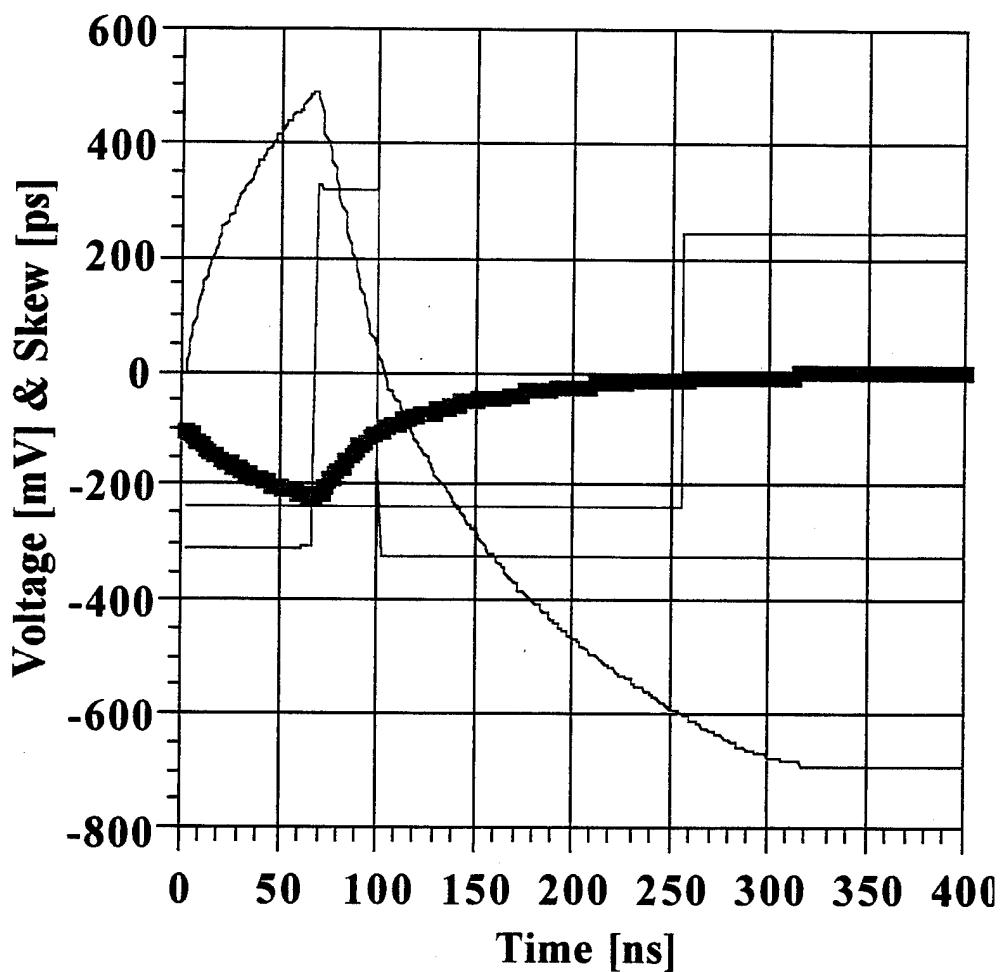

Figure 32. PLL Controller Waveforms.
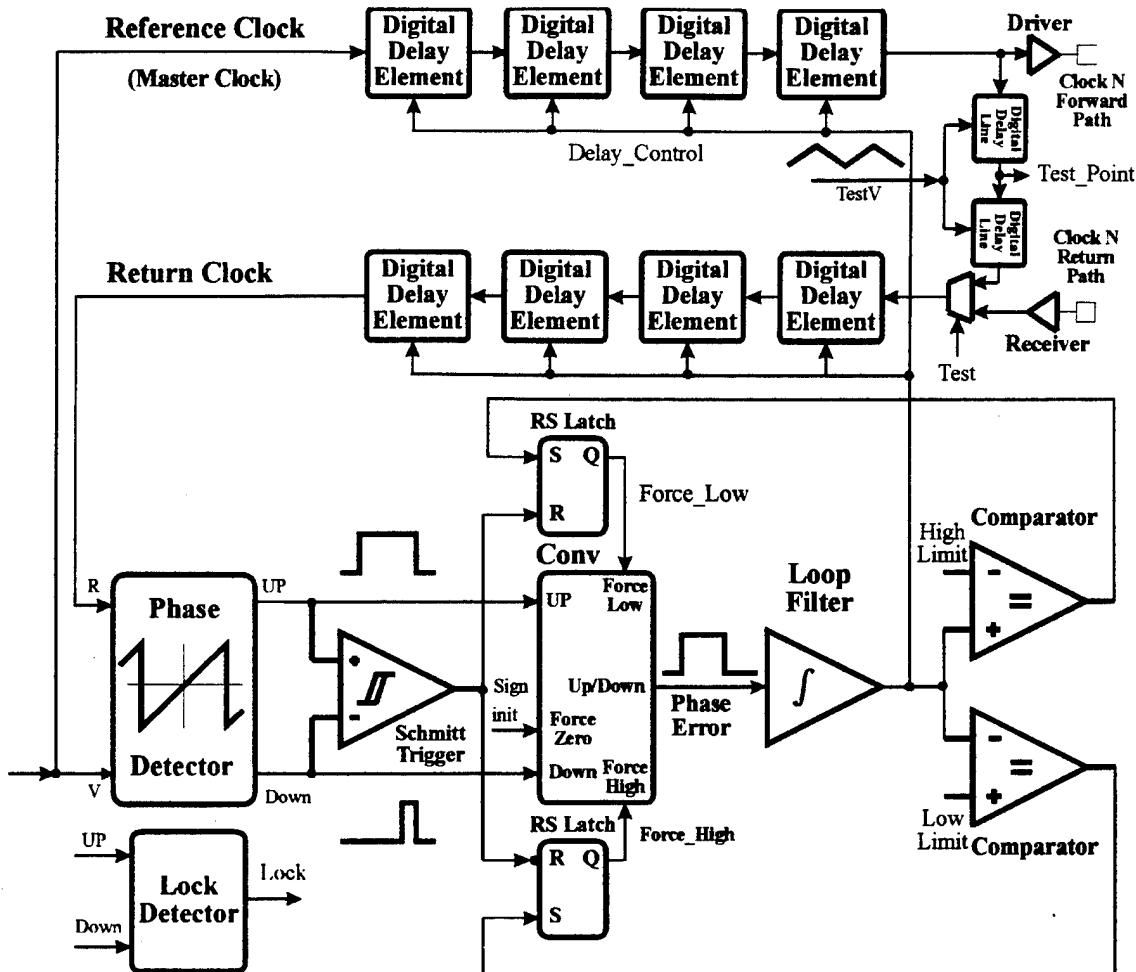
Figure 33. Controller for PLL Clock Loop.

Once the phase detector has changed characteristics the negative feedback loop will drive the PLL into lock (situation 4). Figure 32 shows the PLL controller waveforms and phase error of the PLL for the case were the loop initially latches up. The final phase error is below 5 ps. These PLL waveforms are generated with SPICE. PLLs are difficult to design since PLLs take a very long time to simulate. The transient analysis has to go through hundreds of clock cycles until the steady state is reached. It took 36 hours of CPU time on a Sun10 to generate the traces shown.

VII.3.B.1.Testability

Since the deskew chip will be inserted on an MCM the chip must be fully testable on the wafer for Known Good Die identification. Two additional delays lines have been included in each clock distribution channel to close the clock loop on the chip and simulate slowly varying interconnect delays. This is achieved by applying a slowing varying sawtooth waveform on the TestV input and applying the Test signal. Each channel has a Test_Point signal output to measure skew in test mode. For a more coarse evaluation of a clock channel the phase detector lock signal can also be observed. The lock detector has a window of -15 ps to 15 ps. On the deskew test chip the Test_Point signals of the two clock channels implemented are connected to four phase generators and the ϕ_1 signals are connected to an XOR phase detector. The XOR output signal is connected to an output driver for direct measurements of skew.

Figure 34 shows the layout of the deskew test chip with two clock distribution channels, a system startup controller, and the additional features to increase testability and observability. The deskew test chip contains 1030 HBT devices in an area of 2.6 mm x 3.0 mm and dissipates 2 W.
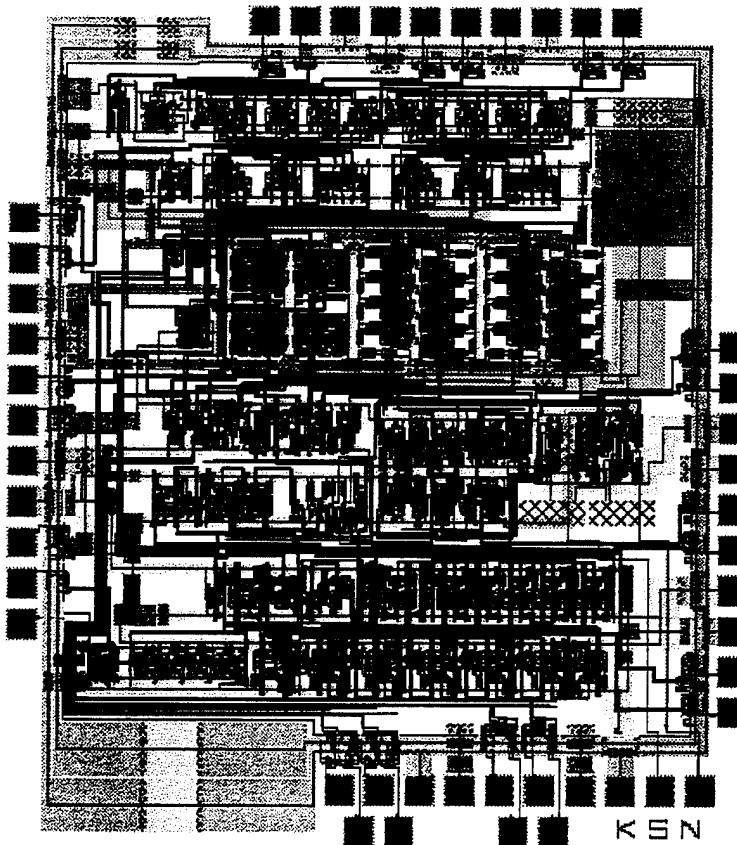

Figure 34. Deskew Test Chip (2.6 mm x 3.0 mm)